Biophysics of V(D)J recombination and Genome Packaging

In singulo studies on RAG, HMGB1, and TFAM

Thesis by

Geoffrey Lovely

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 02, 2014)

Contents

List of Figures

Acknowledgements

I remember sitting on the couch at my apartment in Davis, CA, watching TV with my roommates when I received a phone call from Professor Bil Clemons that stated "we have been trying to get a hold of you for sometime, you have been accepted into the biochemistry and molecular biophysics (BMB) graduate program at Caltech." I was overjoyed. I had assumed I would have an opportunity to work alongside the greatest minds, but these last seven years at Caltech have exceeded my expectations. First, I would like to thank my thesis advisors David Baltimore and Rob Phillips. Our journey began in July 2006, while I was conducting a summer internship at the NIH. I emailed David, informed him of my research interest in exploiting DNA recombination for gene therapy, and asked if his group was still working on zinc finger nucleases. As a sophomore undergraduate I never expected a reply, but the next day I received an email saying "no we are not". So in March 2007 I arrived at Caltech for BMB recruitment, and interviewed with several postdocs in the Baltimore lab who only enhanced my interest in the group. But while I was at BMB recruitment I had the pleasure of meeting our option representative, Rob Phillips, who made it very clear that a biologist equipped with a quantitative skill set has a sixth sense about how things work . After leaving recruitment, I soon realized that getting quantitative was the way to satisfy my interest in exploiting DNA recombination for gene therapy.

In July 2007, I fired off an email to Rob Phillips informing him of my interest to understand DNA recombination with precision to improve the safety of gene therapy. Rob immediately responded and invited me to participate in Caltech's bioengineering bootcamp in September 2007. Bootcamp is an action-packed week of full time experimentation, presentations, discussions, and coding; day one and two gave a brief introduction to microscopy and cloning, and during days three through five we worked on advanced research projects. I was paired with a mathematician, an electrical engineer, and a biophysicist, and had to opportunity to watch a transcription factor carryout acrobatics on a single DNA molecule in real-time. That experiment rocked my world. It became immediately clear to me that all of the processes I learned about in my genetics courses were driven by physical objects or machines. I expressed this excitement to Rob, and while he was supposed to be on sabbatical, he gave me a crash course in statistical mechanics on his chalkboard. After gaining a feeling for the numbers in biology, I proposed two collaborative projects between David and Rob:

the first was to harness synthetic biology and single molecule biophysics to engineer and understand retroviral integration in order to improve their safety as transfer vectors. The second was to use ensemble techniques in David's laboratory and single molecule techniques in Rob's laboratory to study the RAG recombinase, a protein machine discovered in David's laboratory that engineers the genome of white blood cells to make antigen-receptors, which are the heavy artillery they carry to protect use from foreign agents and disease. And like everything in science, nothing is easy, and I hit a roadblock in the first project on retroviral integration, and during that time I was subsequently scooped by three research laboratories pursuing the same idea. I met with David and asked him if he thought it was worth pursuing the project, and he said "it was an engineering problem and it was solved." That was an important lesson because it taught me when it was time to move on. Going forward, I focused on the RAG recombinase, which has ended up being my primary doctoral thesis project. While working on my thesis I had the pleasure of leading bioengineering bootcamp projects in 2008 and 2010, and had the opportunity to work with Professor Jennifer Dionne, JD Bagert, Vanessa Jonsson, Benjamin Gross, and Scott Goodfriend.

I am indebted to my thesis committee members Stephen Mayo, Scott Fraser, and Long Cai, who provided fantastic advice throughout the years. Other Caltech professors who have had significant impact on my personal and professional development include Professors Sarkis Mazmanian, Ellen Rothernberg, David Chan, Pamela Bjorkman, Tom Miller, John Bercaw, Melanie Hunt, Ed Stolper, Joseph Shepherd, Annelia Sargent, John Dabiri, Bil Clemons, Paul Sternberg, Kai Zinn, Judith Campbell, Andre Hoelz, Doug Rees, and Ahmed Zewail. I am thankful to the fantastic group members of the Baltimore and Phillips laboratories, with a special thanks to Joanne Laurence, Alejandro Balazs, Devdoot Majumdar, Yvette Garcia, Michael Bethune, Jesse Bloom, Konstantin Taganov, Mark Boldin, Alex Sigal, Dinesh Rao, Aadel Chaudhuri, Kenneth Yu, Jimmy Zhao, Arnav Mehta, Rachel Galimidi, Katie Miller, Heun Jin Lee, Robert Brewster, James Boedicker, Christoph Haselwandter Franz Weinert, Stephanie Johnson, Tristan Ursell, David Van Valen, David Wu, and Pradeep Ramesh. I am also indebted to my collaborator David Schatz, who took me under his wing as one of his own students.

Caltech certainly would not have been the same without a group of four Caltech graduate student sprinters: Samy Hamdouche, Weston Nichols, Charlie Slominski, Trevor Currie, and one Caltech Professor, Mark Davis. I am also thankful to my community of

friends and colleagues at Caltech, including Ward Walkup, Will Ford, Tony Roy, Philip Romero, Alex Romero, Julian Romero, Nakul Reddy, Alborz Mahdavi, Arseny Vasilyev, Evans Boney, Paul Minor, Adler Dillman, Michael Anaya, Jost Vielmetter, Glenn Garrett, Huu Ngo, and Matt Eichenfield. I am also grateful to the Alfred P. Sloan, Benjamin Rosen, Betty and Gordon Moore, and National Science Foundations for financial support, and to Caltech's Center for Diversity, Graduate Office, and Counseling Center. Lastly, I am and always will be indebted to an incredible woman, my mother Lynn Dell Lovely, who raised four sons, myself, Glenn Lovely, Gary Lovely, and Greg Lovely, and sacrificed her own dreams to work several jobs so that we could have a better life.

Abstract

The recombination-activating gene products, RAG1 and RAG2, initiate V(D)J recombination during lymphocyte development by cleaving DNA adjacent to conserved recombination signal sequences (RSSs). The reaction involves DNA binding, synapsis, and cleavage at two RSSs located on the same DNA molecule and results in the assembly of antigen receptor genes. Since their discovery full-length, RAG1 and RAG2 have been difficult to purify, and core derivatives are shown to be most active when purified from adherent 293-T cells. However, the protein yield from adherent 293-T cells is limited. Here we develop a human suspension cell purification and change the expression vector to boost RAG production 6-fold. We use these purified RAG proteins to investigate V(D)J recombination on a mechanistic single molecule level. As a result, we are able to measure the binding statistics (dwell times and binding energies) of the initial RAG binding events with or without its cofactor high mobility group box protein 1 (HMGB1), and to characterize synapse formation at the single-molecule level yielding insights into the distribution of dwell times in the paired complex and the propensity for cleavage upon forming the synapse. Interestingly, we find the synaptic complex has a mean lifetime of roughly 400s and is highly reversible; only 28% of observed synapses result in cleavage for consensus RSS binding sites. We then go on to investigate HMGB1 further by measuring it compact single DNA molecules. We observed concentration dependent DNA compaction, differential DNA compaction depending on the divalent cation type, and found that at a particular HMGB1 concentration the percentage of DNA compacted is conserved across DNA lengths. Lastly, we investigate another HMGB protein called TFAM, which is essential for packaging the mitochondrial genome. We present crystal structures of TFAM bound to the heavy strand promoter 1 (HSP1) and to nonspecific DNA. We show TFAM dimerization is dispensable for DNA bending and transcriptional activation, but is required for mtDNA compaction. We propose that TFAM dimerization enhances mtDNA compaction by promoting looping of mtDNA.

Chapter 1

Introduction

1.1 The Physical Genome

The genome of an organism contains all the information required for development. The central dogma of biology explains that the genetic information within DNA made of A,T,C, and Gs is transcribed into the language of RNA and translated into the language of proteins. We also know there are on the order of 10^2 - 10^5 protein-encoding genes in a given prokaryotic or eukaryotic genome. Since an organisms genome is made up of only four bases, there must be a method of organizing and preserving the integrity of the genetic information. One level of organization unique to eukaryotes are DNA spools or histones that tightly wrap DNA, and those single histones are organized into eight histones or histone octamers called nucleosomes; the nucleosomes are packaged into solenoids, and then into chromatin composed of euchromatic and heterochromatic regions. Numerous proteins can access chromatin for particular functions e.g. genome packaging, DNA replication, repair, recombination, and transcription. In this thesis we will focus on a type of recombination essential for adaptive immunity, and genome packaging proteins critical for the nuclear and mitochondrial genomes.

1.2 Introduction to V(D)J recombination

1.2.1 Discovery of V(D)J recombination

The specialized lymphoid cells of the jawed-vertebrate adaptive immune system are one of the most fascinating examples of evolution and development. They develop from

a single hematopoietic stem cell and differentiate into adaptive immune cell types called B and T lymphocytes, which can mount a response to foreign molecules called antigens using an almost limitless repertoire of membrane bound antigen receptors [1]. The unique antigen-receptors (ARs) are capable of detecting and evolving to detect virtually any antigen, resultiing in an adaptive immune response. (Note: ARs carried by B or T lymphocytes will be referred to as B cell receptors (BCRs) or T cell receptors (TCRs)). The clonal selection theory, developed in 1957 by Sir Macfarlane Burnet, explains that the adaptive immune response derives from naive B or T lymphocytes carrying BCRs or TCRs that become activated upon binding antigen and then differentiate into mature lymphocytes [2]. Mature B lymphocytes differentiate into plasma cells that secrete their membrane bound BCR as an antibody or immunoglobulin, and mature T lymphocyte TCRs remain membrane bound. Finally, the mature cells become memory cells for sustained immunity.

At the time that the clonal selection theory was proposed, it was clear an adaptive immune response was governed by the specificity and diversity of ARs, but it was unclear how ARs were made. However, Burnet acknowledged that the "[clonal selection] theory requires at some stage in early embryonic development a genetic process for which there is no available precedent," to explain how the repertoire of ARs is assembled [2]. In 1965, while at the Californa Institute of Technology, William J. Dreyer and Jean Claude Bennett first proposed that antibodies were generated by DNA recombination [3] . In 1974, Cesar Milstein showed that the gene segments that make up antibodies, particularly the variable region made of variable (V), diversity (D), and (J) joining gene segments and the constant region were on the same contiguous messenger RNA molecule [4]. However, it was unclear if DNA recombination or RNA splicing joined the two gene regions. In 1976, Susumu Tonegawa conducted experiments that captured the state of the V(D)J genes in the genomic DNA before and after the intervening DNA was cleaved in between, demonstrating that it was DNA recombination and not RNA splicing that assembled ARs [5] . After Tonegawas discovery, he went on to characterize the DNA sequences used as "handles" to assemble ARs that are adjacent to the V, D, and J gene segments, called recombination signal sequences (RSS)s, but it was still unclear what factor was assembling ARs [6-9].

1.2.2 The recombination activating genes, RAG1, and RAG2 carryout V(D)J recombination

The macromolecular machine that carries out V(D)J recombination remained elusive, until two graduate students in the laboratory of David Baltimore, Marjorie Oettinger and David Schatz, discovered two genes aptly named the recombination activating gene-1 and 2, RAG1 and RAG2 by serial transfection of genomic DNA libraries [10,11]. After the discovery of the RAG proteins, researchers found that they were extraordinarily difficult to isolate and study. In the 1990s several laboratories began defining how the RAG proteins work by removing non-essential protein domains, although it was tedious. The N-terminal non-core domain of RAG1 and the C-terminal non-core domain of RAG2 were removed to ease the difficulty of working with the proteins [12-16] . Interestingly, it was later shown that the non-core regions encode a E3 ubiquitin ligase and a plant homeodomain for RAG1 and RAG2, respectively [17, 18] . Core RAG1 (RAG1c) fused to maltose binding protein (MBP) is active and soluble when purified from bacterial, insect, and human cells, and core RAG2 (RAG2c) fused to MBP is only active when purified from insect and human cells, but the most soluble and active RAG1/2c are co-purified from human cells [19]. In Chapter 2 of the thesis I develop a novel approach for purifying RAG1/2c from human cells that improves the protein yield.

The RAG1/2c proteins conduct three major steps at the site of recombination signal sequences (RSSs): site-specific DNA binding, nicking and hairpining. The RSSs consist of a conserved nonamer (ACAAAAACC) and heptamer (CACAGTG) sequence spaced by either 12 or 23bps, aptly named a 12 or 23RSS. After rigorous characterization of RAG1/2c, in 1995 Martin Gellert introduced the first *in vitro* systems for studying the nicking, and hairpining activity of RAG1c and RAG1/2c in the presence of RSSs [20]. In 1996, Gellert, David Schatz, and David Roth independently established evidence for enforcement of the 12/23 rule at the cleavage step of V(D)J recombination in bulk biochemistry assays [21-23]. The 12/23 rule explains that only 12RSS and 23RSS can form hairpins, and as a result assemble functional variable regions [24]. An understanding of RAG1/2c ordered catalysis began to emerge, but the initial binding step to RSSs resisted experimental investigations. That changed in 1997, when a stable RAG1/2c bound to a single 12 or 23RSS was captured, forming the pre-cleavage complex, which is active in cleavage and dependent on high mo-

bility group box protein 1 (HMGB1), especially to achieve equivalent binding to the 23RSS relative to the 12RSS [25]. The existence of a paired complex, which is the intermediate state of RAG1/2c bound to a 12RSS and a 23RSS simultaneously, was inferred from studies on RAG1/2c catalytic activity, but the pre-cleavage paired complex was absent from the literature until its capture on a 12 and 23RSS encoded on different oligonucleotide substrates [26]. The paired complex was shown to require RAG1/2c and HMGB1 for formation on oligonucleotide substrates in the presence of Mg2+, and was robust in the presence of Ca2+ [26]. These biochemical assays have been the stronghold of RAG biochemistry and very useful in understanding RAG proteins on oligonucleotide substrates, but once researchers increased the length of the DNA molecule to 500bps, the assays were not as effect at resolving RAG binding behavior [27]. Current in vitro assays that capture the paired complex with RAG1/2c generally place a 12RSS and 23RSS on two different DNA molecules, typically short oligonucleotides, but in vivo, antigen receptor loci are assembled using RSSs on the same DNA molecule [28]. However, during the assembly of antigen-receptor loci in this correct configuration key mechanistic questions remain unresolved concerning how a diverse immune repertoire is generated by the RAG recombinase. These questions include how long does a RAG complex spend bound to a 12RSS or 23RSS? What is the life-time of a RAG-mediated paired complex formed between a 12 and 23RSS prior to assembling a variable region allele? Is that paired complex reversible or does it always go on to form hairpins, which are the catalytic intermediates formed before a variable region allele is ligated? In chapter 3, we explore the dynamics of various stages of the V(D)J recombination cleavage reaction on single DNA molecules in real-time, allowing visualization and characterization of the action of individual RAG-HMGB1-RSS complexes and paired complex formation using loop lengths equivalent to or 1-3 order of magnitudes less than those found between all V, D and J gene segments. Using a tethered-particle-motion (TPM) assay we we're able to determine single-molecule binding constants for RAG1/2c, the dwell time distributions for the 12RSS and 23RSS protein-bound states, directly observed the reversible paired complex prior to hairpin production, and determined the time-dependence of 12/23 rule-regulated hairpin production at particular concentrations of RAG1/2c and HMGB1. Studies of *in vivo* V(D)J recombination have shown that other factors that effect DNA compaction e.g. modified histones, non-core RAG domains, sequence-specific chromatin bending proteins, promoters-enhancer interactions, non-coding transcription, and the DNA repair machinery,

are essential for the function of V(D)J recombination. We intend to investigate these other factors in the future. In the second part of the thesis we dissect how HMGB1 and another HMG-box protein called transcription factor A mitochondrial (TFAM) compact DNA.

1.3 Introduction to Genome Packaging

1.3.1 High mobility group-box proteins aide in packaging the nuclear genome

The nuclear genome is primarily kept compact by histones, but there are other effectors called "non-histone chromatin associated proteins" that are also involved in packaging the genome. A particular class of these proteins are called high mobility group box proteins (HMGB). Their name is derived from their fast migration during native gel electrophoresis [29]. HMGB proteins are a family of non-specific DNA binding proteins that are conserved in eukaryotes and are the second most ubiquitous DNA binding proteins in the nucleus next to histones. HMGB proteins can also act as chaperones to facilitate DNA binding of tumor suppressors, transcription factors, histones, and the RAG proteins through indirect or direct interactions, and as a result can modulate genome maintenance processes in eukaryotes [30-33].

An HMG box domain is made up of 80 amino-acids [34]. Depending on the HMGB protein they can be found as monomers or two HMGB domains coupled by a flexible linker. The HMGB domains are often equipped with a highly acidic or basic c-terminal tail that has been shown to be critical for both its independent and cooperative functions with other proteins. The dominant HMGB variant, HMGB1, has two HMG box domains coupled by a flexible linker and harbors an acidic tail. It binds, bends and melts the DNA minor groove non-specifically and local deformations or alterations to the geometry of DNA can alter the proteins binding constant for DNA [35, 36]. Single molecule studies using optical tweezers and atomic force microscopy have shown that HMGB1 reduces the persistence length of DNA by introducing a $67^\circ \pm 1.3^\circ$ bend [37, 38]. The persistence length of DNA is defined as the length of DNA in which its end-to-end distance remains constant under thermal forces, which is 50nm or 150bp. The HMGB1-induced kink has been shown to behave as an intermediate between a static kink and a flexible hinge because the linker coupling the two HMG-boxes does not melt the DNA in between [38]. These studies have provided significant

insight into the behavior of HMGB1, but the experimental approach had several caveats, the techniques applied a force, used a DNA molecule of a fixed length, and have ignored the contribution of divalent cations during DNA compaction. In Chapter 4 we investigate how DNA length and divalent cation type effects HMGB1-mediated DNA compaction of single molecules under thermal forces. We find that HMGB1 compacts the same percentage of DNA independent of DNA length, but the divalent cation type can alter the magnitude of DNA compaction at a fixed HMGB1 concentration. Next, we investigate another member of the HMGB family, transcription factor A mitochondrial (TFAM), which regulates the expression and packaging of the mitochondrial genome.

1.3.2 Transcription factor A mitochondrial packages the mitochondrial genome

TFAM is a multi-functional DNA binding protein that is essential for transcriptional activation and mitochondrial DNA (mtDNA) organization [39-42]. As a sequence-specific transcription factor, TFAM plays a central role in production of transcripts from two major mitochondrial promoters, the light strand promoter (LSP) and the heavy strand promoter 1 (HSP1). During *in vitro* transcription assays, normal levels of transcripts from LSP and HSP1 require TFAM [43, 44]. These mitochondrial transcripts encode for 13 essential components of the respiratory chain, and therefore TFAM is necessary for energy production from oxidative phosphorylation [39, 41]. In addition, TFAM is required for maintenance of mtDNA, because truncated transcripts from LSP are used to prime DNA synthesis during mtDNA replication. Consistent with these functions, mice lacking TFAM have impaired mtDNA transcription and an inability to maintain mtDNA, resulting in bioenergetic failure and embryonic lethality [45].

The David Chan group and others have previously solved the crystal structure of TFAM bound to LSP [45, 46]. The TFAM binding site at LSP is 22 base pairs long and composed of two half-sites. TFAM contains two HMG (high mobility group)-box domains (HMG-box A and HMG-box B) that each intercalates into the minor groove of a half-site. Each intercalation contributes to distortion of the DNA, resulting in a dramatic U-turn of the LSP sequence. Between the two HMG-box domains is a helical linker with a positively charged surface that interacts with the negatively charged backbone of DNA. It has been suggested that TFAM may not induce the same bend at HSP1 because C-terminal tail

mutants that cannot bend LSP also do not activate transcription at LSP, but can activate transcription at HSP1, suggesting DNA bending is dispensable and/or TFAM binds in a reverse orientation once bound to HSP1. It has also been suggested that non-specific DNA is not tightly bent because LSP sequence maybe essential. Also, the TFAM/LSP structures revealed an intact homodimerization interface while bound to LSP, and it is unclear what its role is in compacting the mitochrondrial genome. In Chapter 5, we present TFAM/HSP1 and TFAM/non-specific DNA crystal structures, and show that in both cases TFAM imposes a U-turn on the DNA. We confirm TFAM binds HSP1 in the opposite orientation compared to the TFAM/LSP structure, and show the orientation difference underlies the differential DNA bending requirements of the two promoters. Finally, we show that the dimerization interface is not needed for DNA bending or transcriptional activation, but is necessary for full compaction of DNA. These results suggest that TFAM dimerization provides an additional compaction mechanism beyond DNA bending to help organize mtDNA into nucleoids.

1.4 Structure of the thesis

The thesis is organized into two parts. Part I focuses on V(D)J recombination, and Chapters 2 and 3 cover an improvement to RAG protein production and a single molecule analysis of RAG-mediated V(D)J cleavage. Part II is on genome packaging, and Chapters 4 and 5 include a single molecule analysis of HMGB1 and a biophysical analysis of TFAM.

1.5 References

[1] "Immunobiology" Charles Janeway, 7^{th} Edition [2] Burnet M, (1957) Biology and Medicine *Eugen Rev*(3):127-35.

[3] Dreyer WJ and Bennett JC, (1965),The molecular basis of antibody formation: a paradox.*PNAS*54(3):864-9

[4] Milstein C *et al.*(1974) Sequence analysis of immunoglobulin light chain messenger RNA.*Nature*252(5482):354-9.

[5] Hozumi N and Tonegawa S, (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions.*PNAS*73(10):3628-32.

[6] Tonegawa S (1983) Somatic generation of antibody diversity.*Nature*302(5909):575-81.

[7] Steinmetz *et al.* (1980)Two rearranged immunoglobulin kappa light chain genes in one mouse myeloma.*NAR*8(8):1693-707.

[8] Hochtl *et al.* (1982)Recombined flanks of the variable and joining segments of immunoglobulin genes.*PNAS*79(5):1383-7.

[9] Selsing *et al* (1984) Immunoglobulin gene 'remnant' DNA–implications for antibody gene recombination.*NAR*12(10):4229-46.

[10]Schatz DG *et al*, (1989)The V(D)J recombination activating gene, RAG-1. *Cell*59(6):1035-48.

[11] Oettinger M *et al*, (1990)RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.*Science*248(4962):1517-23.

[12] Silver DP *et al.*, (1993) Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. '*PNAS* 90(13):6100-4.

[13] Sadofsky, M *et al.*, (1994) Definition of a core region of RAG-2 that is functional in V(D)J recombination. *NAR* 22(10):1805-9.

[14] Roman CA and Baltimore D, (1996) Genetic evidence that the RAG1 protein directly participates in V(D)J recombination through substrate recognition. *PNAS* 93(6):2333-8

[15] Spanopoulou E *et al.* (1996) The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination *Cell* 87(2):263-76.

[16] Steen, SB *et al.*(1999) Roles of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol Cell Biol* 19(4):3010-7.

[17] Yurchenko V *et al* (2003)The RAG1 N-terminal domain is an E3 ubiquitin ligase.*Genes and Dev*17(5):581-5.

[18] Elkin SK *et al* (2005)A PHD finger motif in the C terminus of RAG2 modulates recombination activity.*J Biol Chem*280(31):28701-10.

[19]Swanson P and Schatz D, (2011) V(D)J recombination: mechanisms of initiation. *Annu Rev Genetics*45:167-202

[20] van Gent DC *et al.* (1995) Initiation of V(D)J recombination in a cell-free system.*Cell* 81(6):925-34.

[21] van Gent DC *et al.* (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination.*Cell*

[22] Steen, SB *et al.*(1999) Roles of the "dispensable" portions of RAG-1 and RAG-2 in

V(D)J recombination. *Mol Cell Biol* 19(4):3010-7.

[23] Eastman QM *et al.*(1996) Initiation of V(D)J recombination in vitro obeying the 12/23 rule.

[24] Kurosawa Y and Tonegawa S (1982) Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA segments.*J Exp Med*55(1):201-18.

[25] D.C. van Gent, *et al.* Stimulation of V(D)J cleavage by High Mobility Group proteins *EMBO J,* 16 (1997), pp. 26652670

[26] Hiom K and Gellert M (1998) Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol Cell* 1(7):1011-9.

[27] Huye LE and Roth DB (2000) Differential requirements for cis and trans V(D)J cleavage: effects of substrate length. *NAR*28(24):4903-11

[28] Swanson PC (2004) The bounty of RAGs: recombination signal complexes and reaction outcomes.*Immunological Reviews*

[29] Johns, E. (ed.) (1982). HMG Chromosomal Proteins. pp. 1251, Academic Press, New York.

[30] Laser, H., *et al.* (2000). A new screen for protein interactions reveals that the Saccharomyces cerevisiae high mobility group proteins Nhp6A/B are involved in the regulation of the GAL1 promoter. *PNAS* 97, 1373213737.

[31] Jayaraman, L., *et al.* (1998). High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev.* 12, 462472.

[32] Kruppa, M., Moir, R. D., Kolodrubetz, D. and Willis, I. M. (2001). Nhp6, an HMG1 protein, functions in SNR6 transcription by RNA polymerase III in S. cerevisiae. Mol. Cell, 7, 309318.

[33] D.C. van Gent, *et al.* Stimulation of V(D)J cleavage by High Mobility Group proteins *EMBO J,* 16 (1997), pp. 26652670

[34] Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility group chromosomal proteins. *Mol. Cell. Biol.* 19, 52375246.

[35] Bustin, M. (2001). Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends *Biochem. Sci.* 26, 152153.

[36] Jaouen S *et al.* (2005) Determinants of specific binding of HMGB1 protein to hemicatenated DNA loops.*J. Mol Biol*353(4):822-37

[37] McCauley,M.J., *et al* . (2007) HMGB binding to DNA: single and double box motifs.

J. Mol. Biol., 374, 9931004.

[38] Zhang, J., *et al* (2009). Mechanism of DNA flexibility enhancement by HMGB proteins. *NAR* 37, 11071114.

[39]. Bonawitz, N.D., *et al* (2006). Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Mol.Cell* 24, 813-825.

[40]. Campbell, C.T., *et al* (2012). Mitochondrial transcription factor A regulates mitochondrial transcription initiation, DNA packaging, and genome copy number. *Biochim Biophys Acta* 1819, 921-9.

[41] Falkenberg, M., *et al* (2007). DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem* 76, 679-99.

[42.] Rubio-Cosials, A. and M. Sola (2013). U-turn DNA bending by human mitochondrial transcription factor A. *Curr Opin Struct Biol* 23, 116-24.

[43]. Falkenberg, M., *et al* (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat.Genet.* 31, 289-294.

[44]. Shi, Y., *et al* (2012). Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *PNAS*109, 16510-16515.

[45]. Larsson, N.G., *et al* (1998). Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat Genet* 18, 231-6.

[46]. Ngo, H.B.,*et al* (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat.Struct.Mol.Biol.* 18, 1290-1296.

Part I

V(D)J Recombination

Chapter 2

Improved production of the RAG proteins from human cells.

2.1 Abstract

The recombination activating genes, RAG 1 and RAG2, assemble the antigen-receptor repertoire. Since their discovery full-length RAG1 and RAG2 have been difficult to purify, and core derivatives were shown to be most active when co-purified from adherent 293-T cells using pcDNA1 expression vectors. However, the protein yield from adherent 293-T cells transfected with pcDNA1 is limited. This bottleneck in protein production restricts the structural and biophysical characterization of proteins. Here, we develop a human suspension cell purification and change the expression vector to boost RAG production. We compare RAG protein expression from pcDNA1 or pTT5 vectors with the Simian virus-40 (SV40) or Epstein-Barr virus (EBV) origin in 293-T cells carrying the SV40 large T-antigen and 293-E cells harboring the EBV nuclear antigen. We show that 293-E and 293-T cells transfected with pcDNA1 yield an equivalent amount of RAG protein, but we observe a 6-fold increase in RAG production using pTT5 compared to pcDNA1 in 293-E cells. We also show the suspension purification does not alter the oligomeric state or function of the RAG proteins. Our large-scale purification reduces a significant bottleneck in RAG protein production that should lower the barrier-to-entry for structural and biophysical studies.

2.2 Introduction

Variable (V), diversity (D), and joining (J) segment recombination is responsible for assembling all of the antigen-receptors during lymphoid cell development. During V(D)J

recombination, fragments of alleles interspersed on particular chromosomes are assembled into functional alleles by RAG1/2 [1]. During V, D, and J assembly, D to J rearrangement precedes V to DJ rearrangement. The recombined alleles subsequently become genes that encode antigen-receptors called immunoglobulin (Igs) or antibodies in B lymphocytes or T cell receptors (TCR) in T lymphocytes. The RAG proteins recognize Igs and TCR gene segments though recombination signal sequences (RSS) adjacent to particular V, D, and J gene segments that are composed of a conserved nonamer (ACAAAAACC), a spacer (12 or 23bp), and a conserved heptamer (CACAGTG) [1]. Most of what we know about V(D)J recombination is derived from studies in cells, because biochemical work has been extraordinarily laborious because of difficulties in protein expression and purification. The first attempts to purify full-length RAG (FLRAG1/2) proteins showed that the proteins were inactive and insoluble. Although the process was tedious, several laboratories removed non-essential protein domains from FLRAG1 and FLRAG2. The N-terminal non-core domain of RAG1 and the C-terminal non-core domain of RAG2 were removed [2-8]. It was later shown that the non-core domains encode a E3 ubiquitin ligase and a plant homeodomain for FLRAG1 and FLRAG2, respectively [9, 10]. These core RAG1 and RAG2 protein domains, which will be referred to as RAG1c and RAG2c (RAG1/2c), conduct three major steps at the site of RSSs sequence specific DNA binding, nicking, and hairpin formation [11-13].

Once RAG1c and RAG2c are expressed, they usually require covalently coupled proteins to increase their solubility [1]. The maltose binding protein (MBP) is used most frequently. Active MBP-RAG1c can be successfully purified from bacteria, insect, and mammalian cells; however, active MBP-RAG2c is only recovered from insect and mammalian cells [1]. Currently the most active RAG1/2c proteins recovered are co-expressed in adherent 293-T cells at a 1:1 ratio, using pcDNA1 expression vectors. pcDNA1 carries the simian virus 40 (SV40) origin of replication, and 293-T cells harbor the SV40 large T antigen, which replicates SV40 expression vectors. Although co-expressed RAG1/2c drastically improve the quality of RAG protein activity, the protein yield from pcDNA1 is minimal. To ameliorate this issue, we increased RAG protein production by making two modifications to the current purifications. First, we grew the cells in suspension, which increased the absolute number of cells grown in culture and reduced the amount of time to grow enough cells for transfection. Second, we changed the expression vector from pcDNA1 to pTT5. pTT5 carries the Epstein-Barr virus (EBV) origin of replication, and when expressed in 293-E

cells that have the EBV nuclear antigen (EBNA1), the pTT5 vector undergoes episomal replication, which increases the number vectors driving RAG expression.

2.3 Results

2.3.1 Increasing RAG protein production

We cultured 100mL of suspension 293-T or 293-E cells and co-transfected the cells using a lipofection reagent with pcDNA1 or pTT5 vectors carrying MBP-RAG1c and MBP-RAG2c. After 48 hours we counted the cells. The 293-T and 293-E inital cell density was 1.0 x 10^6 cells/mL. During the 48 hour transfection the 293-T cells transfected with pcDNA1 or pTT5 grew to 2.49 x 10^6 $\pm$ 0.33 x 10^6 cells/mL or 2.10 x 10^6 $\pm$ 0.14 x 10^6 cells/mL, respectively. The 293-E cells transfected with pcDNA1 or pTT5 grew to 1.76 x 10^6 $\pm$ 0.31 x 10^6 cells/mL or 1.85 x 10^6 $\pm$ 0.45 x 10^6 cells/mL, respectively. The final cell density increased 2-fold relative to the initial 293-T and 293-E cell density (Fig 2.1a). Next, we harvested and centrifuged the cells, added lysis buffer to each pellet, and Dounce homogenized the sample. Then we ultracentrifuged the homogenate, and added the cleared lysate to amylose resin. We poured the cleared lysate/amylose resin mix onto a gravity flow column, collected the flow-through, washed the resin twice, and eluted with maltose (Fig 2.1b,c). The procedure was repeated, and RAG protein concentration and the fraction of contaminating DNA from elution fractions were determined using a nanodrop (see Table 1). We computed the absolute amount of protein produced by a single 293-T or 293-E cell, using the cellular density (CD) after 48 hours, culture volume (CV), RAG concentration ([RAG]), and elution volume (EV) of 300μL. We determined the absolute amount of protein purified from a single cell for each independent purification (shown in Table 1) using the equation

$$RAG\ production\ per\ cell = \frac{((EV)([RAG]))}{((CD)(CV))} \tag{2.1}$$

We found that pTT5 produced more RAG protein than pcDNA1 in both 293-E and 293-T cells; however, 2.41$\pm$0.28pg was recovered per 293-E cell ,which is 1.6-fold greater than the 1.32$\pm$0.35 pg recovered per 293-T cell.(Fig 2.1d). Interestingly, pcDNA1 produced 0.36$\pm$0.05 pg per 293-E cell, which is 6.2-fold less than pTT5 (Fig 2.1d). Therefore pTT5 and 293-E cells are optimal for RAG protein production. After improving RAG production,

we investigated the oligomeric state and function of the proteins to confirm they were unaffected by higher protein expression.

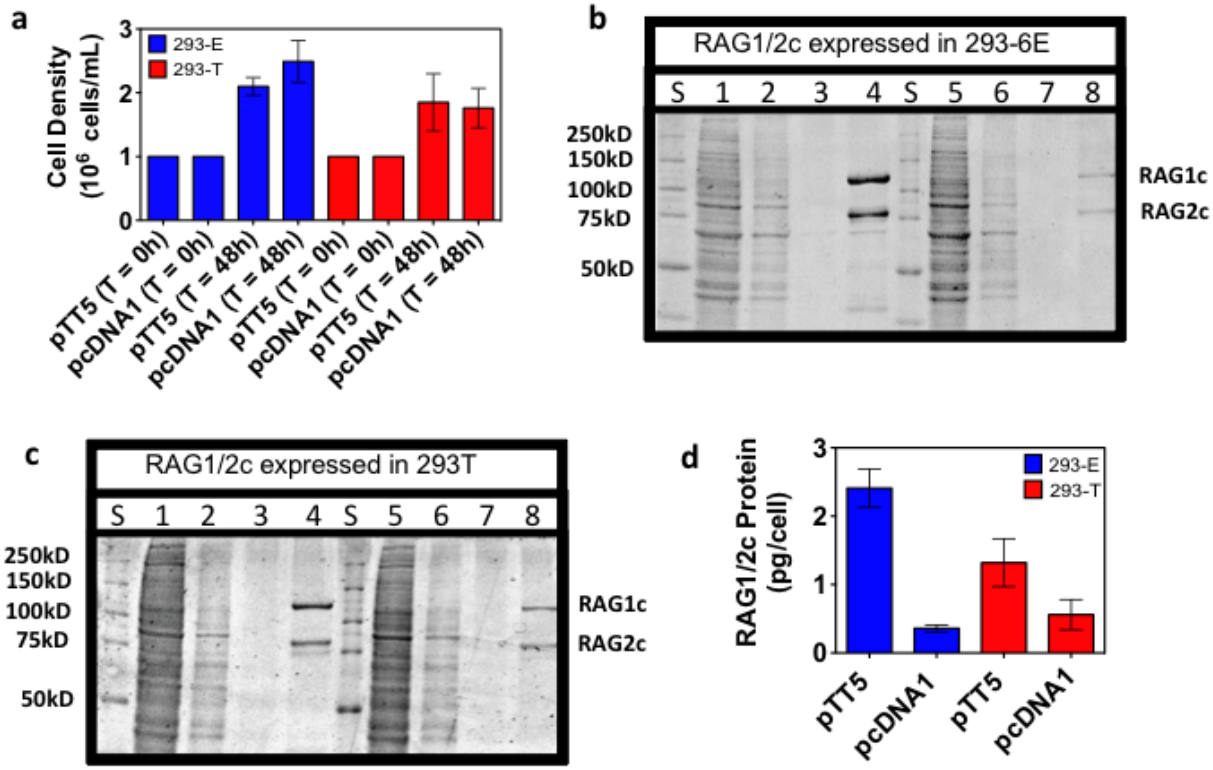

Figure 2.1: Expression and purification of RAG1/2c from suspension 293-T and 293-E cells. a) 100mL 293-E (blue) or 293-T (red) cells were diluted to 1.0 x 10^6 cells/mL before (T= 0h) and after (T= 48h) transfection the cellular density was determined for cells transfected with either pTT5 or pcDNA1 b) and c) SDS-PAGE of MBP-RAG1/2c expressed in 293-E or 293-T cells respectively, using pTT5 (lanes 1-4) or pcDNA1 (lanes 5-8). The lanes represent different stages of the protein purification, the flow through (lanes 1 and 5), wash 1 (lanes 2 and 6), wash 2 (lanes 3 and 7) and the elution: lanes 4 and 8]. Lane S are proteins standards ranging from 250kD to 50kD. RAG1c and RAG2c labels are where RAG proteins are expected to run on the gel. d) Total protein produced per 293-E cell (blue) or 293-T (red) using pTT5 or pcDNA1. We used the equation RAG production per cell = (elution volume x [RAG])/ (cell density x culture volume) to determine values.

2.3.2 SEC-MALS and functional analysis of suspension purified RAG recombinase

We measured the molecular weight of MBP-RAG1/2c to confirm that the increase in protein production did not alter the oligomeric state of MBP-RAG1/2c purified from high-salt, which is predominately a dimer of RAG1 (120kD) and one RAG2 (80kD) with a predicted

Table 1: Cell density, RAG1/2c elution and DNA contamination

	293-T Cells: Cell Density (10^6 cells/ml)	293-T Cells: A280 (mg/ml)/ A260:A280 ratio*	293-E Cells: Cell Density (10^6 cells/ml)	293-E Cells A280(mg/ml)/ A260:A280 ratio*
1) pTT5-RAG	2.2	1.22/0.76	2.2	1.62/0.74
2) pTT5-RAG (D)	1.5	0.52/0.87	2.0	1.74/0.70
3) pcDNA1-RAG	2.0	0.49/0.86	2.3	0.25/1.06
4) pcDNA1-RAG (D)	1.5	0.20/1.08	2.7	0.36/1.05

*260/280 of 0.57 = 100% Protein/0% DNA contamination
260/280 of 1.06 = 95% Protein/5% DNA contamination
(D) = Duplicate

molecular mass of 320kD [14]. We carried out size exclusion chromatography-multiangle light-scattering (SEC-MALS) on RAG1/2c purified using pTT5 expression vectors and 293-E cells. We measured an average molecular weight of 324kD ± 19.79 kD, which is within error to the predicted molar mass (Fig 2.2a). After showing that the RAG oligomeric state was unaltered, we performed functional assays to confirm the proteins behave the way the RAG proteins do in the literature. We conducted an EMSA using long (500bp) DNA molecules, RAG1/2c, competitor DNA, and its co-factor high mobility group box-1 (HMGB1), and detected discrete RAG:HMGB1:DNA complexes only in the presence of RSSs (lane 8) (Fig 2.2b). We also conducted two *in vitro* cleavage assays using long DNA molecules and oligonucleotides, and showed that the suspension RAG proteins obey the 12/23 rule during HMGB1-dependent hairpin production (Fig 2.2c and d).

2.4 Discussion

Our study provides a new method for purifying large amounts of active RAG1/2c from human cells. Previous methods showed that 293-T cells were optimal for producing the most active RAG proteins, but culturing the cells in an adherent configuration limited protein production [15]. We boosted RAG production by changing the 293-T to 293-E cells and the

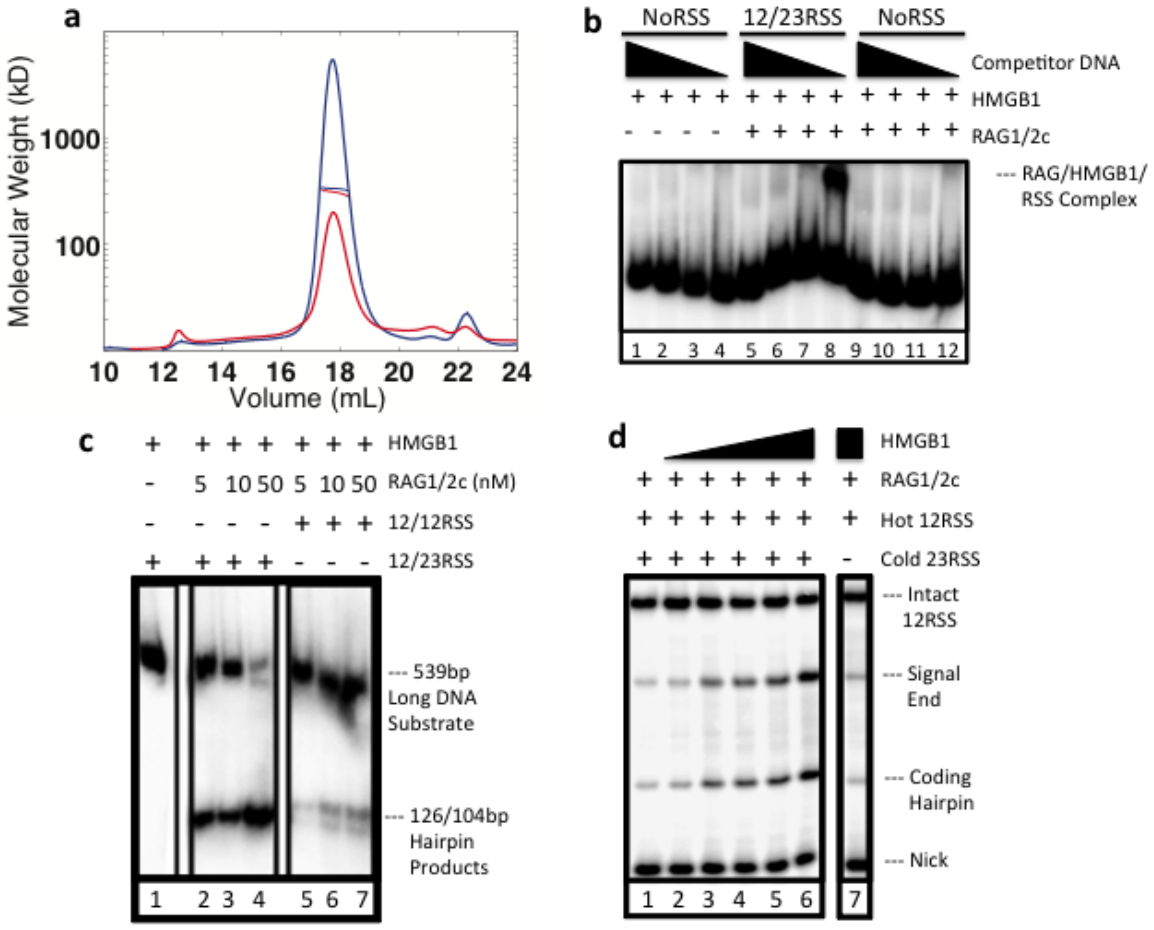

Figure 2.2: Size exclusion chromatography-multiangle light-scattering (SEC-MALS) and functional analysis of RAG1/2c. a) SEC-MALS distributions and calculated molecular mass for two independent (red and blue) purifications from 293-E cells. b) We conducted an EMSA using 50nM RAG1/2c, 225nM HMGB1, decreasing amounts of single-stranded salmon sperm competitor DNA (300ng, 200ng, 100ng, and 50ng) and target DNA with (12/23RSS) or without (NoRSS) RSSs. HMGB1 alone was added to the NoRSS (lanes 1-4) and RAG proteins and HMGB1 with (12/23RSS: lanes 5-8) or without (NoRSS: 9-12) RSSs. Binding reactions were conducted at 4 °C. *In vitro* cleavage assays were conducted using c) long DNAs in the presence of 225nM HMGB1,increasing amounts RAG1/2c [5, 10 and 50nM], and 300ng of radiolabeled 12/23RSS or 12/12RSS DNA substrates. The cleavage reactions were fractioned on a 4% non-denaturing gel. d) Oligonucleotide substrates were also used with 50nM RAG1/2c, increasing amounts of HMGB1 [5, 10, 25, 50 112nM], 1nM radiolabeled 12RSS and 4nM 23RSS.

cellular growth conditions from adherent to suspension. Now we can purify RAG proteins from many liters of 293-E suspension cells grown in parallel. In addition, we changed the vehicle for expressing the RAG proteins from pcDNA1 to pTT5 to further increase protein production. We showed that we could produce 2.41 ±0.28pg of RAG proteins per 293-E cell using pTT5, six times greater than pcDNA1. The difference in RAG protein production between pTT5 and pcDNA1 in 293-E cells is caused by episomal replication of pTT5 by EBNA1 within 293-E cells. The combination of pTT5 vector's EBV origin and the EBNA1 antigen in 293-E cells increased RAG expression. We also showed the final cellular density of 293-E cells after transfection was 2.1 x 10^6 ± 0.14 x 10^6 cells/mL, giving a RAG protein yield from 100mL of 0.5mg. We have scaled up the purification to 1L and recovered 5mg of RAG protein (data not shown) as predicted from the small-scale purification. The RAG proteins are now positioned for precise structural and biophysical characterization.

2.5 Methods

2.5.1 Expression Vectors

An MBP fusion to RAG1c or 2c was cloned into the pTT5 expression vector, from pcDNA1-MBP-RAG1c or RAG2c constructs, which were a kind gift from Professor David Schatz via Professor Patrick Swanson. The pTT5 vector was obtained from the National Research Council Canada (NRCC)

2.5.2 Suspension cell culture

Suspension HEK293-E from the NRCC or HEK293-T cells from Caltech's Protein Expression Center (PEC)were cultured in Freestyle 293 media (Gibco); however, there are slight modifications in how each cell type was maintained. The HEK293-E cells were split to a minimal dilution of 1 x 10^6 cells/mL. During routine maintenance the cells are grown to 2 x 10^6 cells/ml which takes 24h and then split back down to 1 x 10^6 cells/ml; however, before a transfection the cells are grown to 3 x 10^6 cells/mL and diluted to 1 x 10^6 cells/mL. The HEK293-T cells were cultured in freestyle 293 media and 0.1% Pluoronic (Gibco) to prevent clumping. Then they were split to a minimal dilution of 0.8 x 10^6 cells/mL and were grown to a maximal density of 2 x 10^6 cells/mL. Prior to transfection the cells were diluted to 1 x 10^6 cells/mL.

2.5.3 Protein expression and purification

RAG proteins were purified from suspension 293-E cells using CMV driven pcDNA-MBP-RAG1/2c or pTT5-MBP-RAG1/2c expression vectors harboring SV40 or EBV origins, respectively. First we transformed pTT5-MBP-RAG1 or 2 harboring ampicillin resistance in Z-competent cells (Zymoresearch) and plated on LB-agar containing 50ug/mL carbenicillin (Sigma-Aldrich). A single colony was selected and a large scale plasmid purification was performed using the PowerPrep HP Plasmid Maxiprep kit (Origene). Next, 100mL of suspension 293-E cells were grown to a cell density of 1 x 10^6 cells/mL, prior to transfection in Gibco freestyle 293 expression medium (Life Technologies). 100m: of 293-E cells were transfected with 45μg of pTT5-MBP-RAG1, 55μg of pTT5-MBP-RAG2, and 100μL of BioT (Bioland) brought to a final volume of 3.5mL in 1X PBS (Note: Mix DNA and 1X PBS first). The transient transfection proceeds for 48h and the cells were harvested at 3,000 rpm (700g) for 15 min. Next the cells pelleted were resuspended in 3mL of lysis Buffer A (10mM sodium phosphate pH 7.4, 500mM NaCl, 1mM DTT, 0.25% Tween-20, 10μg/ml aprotinin, 10μg/ml leupeptin, 10μg/ml pepstatin, and 100μM phenylmethanesulfonylfluoride (PMSF)) and Dounced on ice with 20 strokes. The cell lysate was centrifuged at 86,000g for 40 min using the SW28 rotor (Beckman). The supernatant was then added to 300μL of amylose resin (New England Biolabs) and inverted for 1h 30 min. The amylose resin/supernatant were poured into a bio-spin chromatography column (Bio-Rad). The amylose resin complex was washed with 2ml of Buffer A, followed by 1mL of Buffer B (10mM sodium phosphate pH 7.4, 500mM NaCl, 1mM DTT), 75μL of elution buffer (10mM sodium phosphate pH 7.4, 500mM NaCl, 1mM DTT, and 10mM maltose) to displace column void volume, and finally 300μL of elution buffer. The protein was dialyzed for 3h in dialysis buffer (25mM Tris-HCl (pH: 8.0), 150mM KCl, 2mM DTT, and 10% glycerol) using a 0.1 to 0.5mL slide-a-lyzer (Pierce), nap-frozen and stored at $-80\,^{\circ}$C.

2.5.4 SDS-PAGE

10% Stacking gels were used in all the SDS-PAGE gels, and were stained with sypro orange (Life Technologies) for 10 min and visualized with a transilluminator.

2.5.5 Electrophoretic mobility shift assay

The gel-shift assay was performed as previously described [16]. First, 125nM RAG1/2c and 560nM HMGB1 were preassembled in an 8μL reaction in dilute dialysis buffer (DDB: 25mM Tris-HCl (pH: 8.0), 150mM KCl, 2mM DTT) and 0.1% glycerol for 5 min. Then, 12μL of EMSA reaction buffer (ERB: 25mM Tris-HCl (pH: 7.4), 65mM potassium acetate, 5mM MgCl2, 100μg/ml acetylated BSA (Sigma-Aldrich), and 10% DMSO) containing 65nM of 539bp NoRSS or 12/23RSS DNA substrate was added to RAG1/2c and HMGB1, bringing the DDB and ERB together to form the RAG reaction buffer (RRB) and the reaction volume to 20μL with 50nM RAG1/2c, 225nM HMGB1, and 40nM NoRSS or 12/23RSS. The reaction was incubated at 4 °C for 15 min, then decreasing amounts of competitor DNA 300ng, 200ng, 100ng, and 50ng were added and the reaction incubated for another 10 min at 4 °C. The reactions were fractionated on a 4% 29.1 (Bio-Rad) 1/2X TBE polyacrylamide gel for 1h and 30 min at 250V using a model V16, and visualized with a phosphoimager.

2.5.6 *In vitro* cleavage assay

The oligonucleotide RAG cleavage assay detects nicked and hairpin products, and was performed as previously described [15]. Briefly, 50nM of RAG1/2c and increasing amounts of HMGB1 (10nM, 25nM, 50nM, and 112nM) were assembled as described in the EMSA in the presence of 5nM 12RSS and 10nM 23RSS, with one reaction containing 112nM HMGB1 but no 23RSS, and incubated at 37 °C for 1hr. The reaction was stopped with 95% formamide and 10mM EDTA and incubated at 95 °C for 2 min. The reaction products were fractionated on on 15% 19:1 polyacrylamide sequencing gel (National Diagnostics) and visualized with a phosphoimager.

2.5.7 Size exclusion chromatography and multiangle light-scattering

Size exclusion chromatography coupled to multiangle light-scattering (SEC-MALS) was carried out using a Varian HPLC with three in-line detectors, as previously described [17]. Detectors were a Varian Prostar 345 UV detector followed by a DAWN HELEOS-MALS instrument (Wyatt Technologies) and an Optilab rEX (Wyatt Technologies). Samples were run over a Shodex KW803 column at 0.5 mL/min in 25 mM Tris pH 7.5, 150 mM KCl, 2mM DTT, and 10% Glycerol at 25 °C. Data was processed using the protein conjugate

analysis implemented in Astra 5.3.4.16. Protein molecular weight values are the average of two injections.

2.6 References

[1]Swanson P and Schatz D, (2011) V(D)J recombination: mechanisms of initiation. *Annu Rev Genetics*45:167-202

[2] Oettinger M *et al*, (1990)RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.*Science*248(4962):1517-23.

[3] Silver DP *et al.*, (1993) Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. '*PNAS* 90(13):6100-4.

[4] Sadofsky, M *et al.*, (1994) Definition of a core region of RAG-2 that is functional in V(D)J recombination. *NAR* 22(10):1805-9.

[5] Roman CA and Baltimore D, (1996) Genetic evidence that the RAG1 protein directly participates in V(D)J recombination through substrate recognition. *PNAS* 93(6):2333-8

[6] Spanopoulou E *et al.* (1996) The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination *Cell* 87(2):263-76.

[7] Steen, SB *et al.*(1999) Roles of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol Cell Biol* 19(4):3010-7.

[8] Sekiguchi JA *et al.* (2001) Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol Cell* 8(6):1383-90.

[9] Yurchenko V *et al* (2003)The RAG1 N-terminal domain is an E3 ubiquitin ligase.*Genes and Dev*17(5):581-5.

[10] Elkin SK *et al* (2005)A PHD finger motif in the C terminus of RAG2 modulates recombination activity.*J Biol Chem*280(31):28701-10.

[11 van Gent DC *et al.* (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination.

[12] Steen, SB *et al.*(1999) Roles of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol Cell Biol* 19(4):3010-7.

[13]Eastman QM *et al.*(1996) Initiation of V(D)J recombination in vitro obeying the 12/23 rule.

[14Raval P *et al.* (2008) Evidence for Ku70/Ku80 association with full-length RAG1.*NAR*

36(6):2060-72.

[15] Bergeron S *et al.* (2006) RAG and HMGB1 proteins: purification and biochemical analysis of recombination signal complexes. *Methods Enzmol* 408:511-28.

[16] Kumar S and Swanson PC (2009) Full-length RAG1 promotes contact with coding and intersignal sequences in RAG protein complexes bound to recombination signals paired in cis. *NAR* 37(7):2211-26.

[17] Gandhi CS *et al.*,(2011) OCAM: a new tool for studying the oligomeric diversity of MscL channels.*Protein Sci* 20(2):313-26

Chapter 3

Single-molecule analysis of RAG-mediated V(D)J cleavage

3.1 Abstract

The recombination-activating gene products, RAG1 and RAG2, initiate V(D)J recombination during lymphocyte development by cleaving DNA adjacent to conserved recombination signal sequences (RSSs). The reaction involves DNA binding, synapsis, and cleavage at two RSSs located on the same DNA molecule and results in the assembly of antigen receptor genes. We have developed single-molecule assays to examine RSS binding by RAG1/2c and their co-factor HMGB1 as they proceed through the steps of this reaction. This assay allows us to observe, in real time, the individual molecular events of RAG-mediated cleavage. As a result, we are able to measure the binding statistics (dwell times and binding energies) of the initial RAG binding events and to characterize synapse formation at the single-molecule level yielding insights into the distribution of dwell times in the paired complex and the propensity for cleavage upon forming the synapse. Interestingly, we find the synaptic complex has a mean lifetime of roughly 400s and is highly reversible; only 28% of observed synapses result in cleavage for consensus RSS binding sites

3.2 Introduction

V(D)J recombination is responsible for assembling the variable regions of antigen receptor genes during B and T lymphocyte development. During V(D)J recombination, fragments of V (variable), D (diversity) and J (joining) segments located together on particular chromosomes are rearranged into functional VDJ or VJ alleles that are the specificity deter-

minants for B cell receptors, or immunoglobulins (Ig), and T cell receptors [1]. Adjacent to the V, D and J gene segments are recombination signal sequences (RSS) with a conserved heptamer (consensus 5'-CACAGTG-3') and nonamer (consensus 5'-ACAAAAACC-3') separated by a spacer of 12 or 23bp (referred to as 12RSS and 23RSS, respectively). Efficient recombination requires one 12RSS and one 23RSS, a constraint known as the 12/23 rule [1].

The recombination activating genes, RAG1 and RAG2, encode proteins that carry out V(D)J recombination by bringing a 12RSS and a 23RSS together into a paired (or synaptic) complex, nicking the RSS sites adjacent to the heptamer, and converting the nicks into a double strand break and a hairpin at the coding end. The hairpin is created by nucleophilic attack on the opposing strand by the 3 hydroxyl group at the nick in a transesterification reaction. After RAG1/2 forms hairpins, it recruits the non-homologous end joining (NHEJ) machinery to repair the ends [1, 2]. Since their discovery, the full-length RAG1 and RAG2 proteins have proven difficult to isolate and study in vitro [1]. However, core domains (referred to as RAG1c and RAG2c) have been identified by removing a large region from the N terminus of RAG1 (which includes an E3 ubiquitin ligase domain) and a large region from the C terminus of RAG2 (which includes a plant homeodomain) [3, 4]. These core proteins have been shown to tetramerize to form RAG1/2c, which retains RSS binding, nicking and hairpin formation activities [5]. High mobility group-box protein 1 (HMGB1) acts as a co-factor and increases RAG1/2c affinity for the 23RSS [6]. HMGB1 is required for paired complex formation and efficient conversion of RAG-mediated nicks into hairpins [7-10].

Current in vitro assays that capture the paired complex generally place a 12RSS and 23RSS on two different DNA molecules, typically short oligonucleotides [1]. However, in vivo, antigen receptor loci are assembled using RSSs on the same DNA molecule [11]. One prior study captured RAG1/2c and HMGB1 bound to DNA with a 12RSS and 23RSS on the same substrate using standard bulk assays [12]. Here, we explore the dynamics of various stages of the V(D)J recombination cleavage reaction on single DNA molecules in real-time, allowing visualization and characterization of the action of individual RAG-HMGB1-RSS complexes and paired complex formation using loop lengths equivalent to those found in the immunoglobulin heavy chain (IgH) locus between particular D and J gene segments in B cells [13] . Using a tethered-particle-motion (TPM) assay [14] we observed RSS-dependent

apparent shortening of DNA in the presence of RAG1/2c and were able to exploit this shortening as a signature of the RAG1/2c-RSS bound state. This has allowed us to determine single-molecule binding constants for RAG1/2c using our measurements in conjunction with a statistical mechanical model of RAG binding. These same measurements are also used to determine the dwell time distributions for the protein-bound states. Furthermore, this same substrate length readout is used for direct detection of the paired complex prior to hairpin production. These complexes have one of two eventual outcomes: they lead to cleavage of the DNA allowing the bead to drift away or the paired complex reversibly dissociates and leaves the DNA template intact. Finally, we use bead loss upon cleavage to quantify the time-dependence of cleavage on the concentrations of RAG1/2c and HMGB1.

3.3 Results

3.3.1 Detection of single RAG-RSS complexes in real-time

We have developed a single-molecule assay to capture RAG1/2c-12RSS or RAG1/2c-23RSS complexes using a technique called tethered particle motion (TPM), previously used in studies on genome dynamics [15-17], that exploits the relation between DNA length and root-mean square (RMS) motion of the bead to detect length changes in DNA tethered beads (Figure 3.2) [14]. Figure 1A introduces the various microscopic states that we seek to observe in the context of our single-molecule experiments. We explore the properties of RAG1/2c-DNA interaction for molecules harboring no RSS, only the 12RSS, only the 23RSS, and both 12RSS and 23RSS together using co-expressed and co-purified maltose binding protein (MBP)-tagged RAG1/2c proteins from a mammalian cell line in a buffer that includes 2.5mM Mg2+ and 5% DMSO.

To address the question of how the RAG proteins interact with different constellations of RSS sites, we added RAG1/2c at concentrations between 1nM and 50nM to the various DNA constructs. We found that, even at a saturating concentration of 50nM RAG1/2c at which RAG1/2c is stably bound to DNA in the presence of RSS binding sites, without an RSS binding site there was no measurable change in the mean RMS motion of the bead, indicating that if RAG1/2c does bind nonspecifically to DNA that the binding does not produce a measurable change in the effective length of the DNA (Figure 3.4(D,J)).

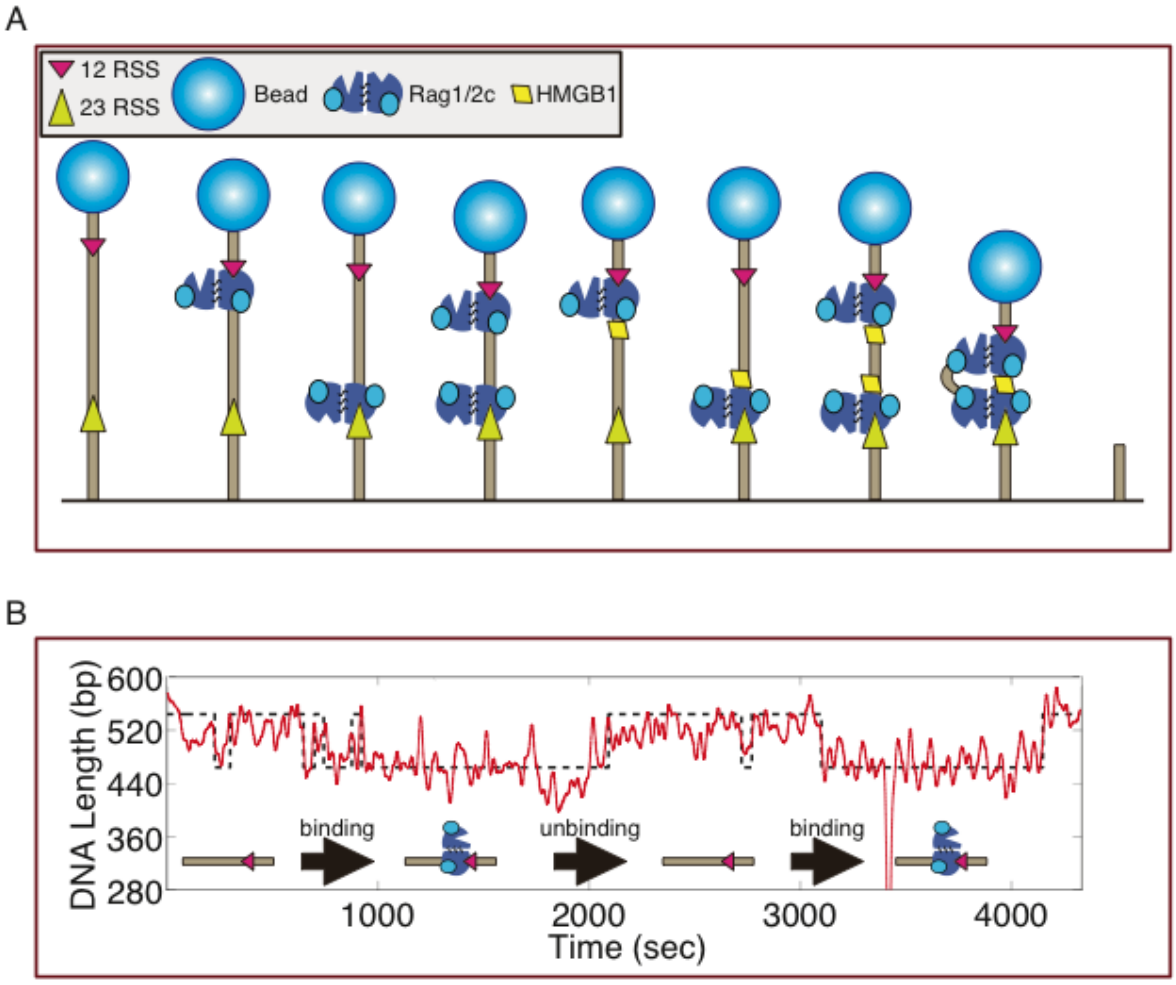

Figure 3.1: Schematic of states associated with a single V(D)J recombination event. (A) The DNA molecule (tan) has a 12RSS (maroon) and a 23RSS (yellow). One RAG1/2c (cyan) complex can bind either a 12 or 23RSS, and two RAG1/2c complexes can bind two RSSs. One RAG1/2c and an unknown number (X) of HMGB1 (yellow rhombus) molecules can bind a 12 or 23RSS, two RAG1/2c and X HMGB1 can bind two RSSs. A final HMGB1-dependent state brings 12 and 23RSS together forming a paired complex and a loop in the DNA which will shorten the DNA molecule. There is some uncertainty as to how many RAG1/2c and HMGB1 molecules are required to form this complex. Assembly of the RAG1/2c-HMGB1 complex can ultimately result in severing of the DNA (hairpin formation). In this study we used a technique called tethered-particle motion that can provide a direct window into the behavior of DNA binding proteins that shorten DNA. We tethered one of these DNA molecules with RSSs to a surface and coupled a bead (light-blue) to it and watched the action of RAG1/2c with or without HMGB1. (B) Example trajectory of a bead in the presence of 5nM RAG1/2c as it transitions between unbound states (characterized by longer apparent lengths) and bound states (characterized by shorter apparent lengths).

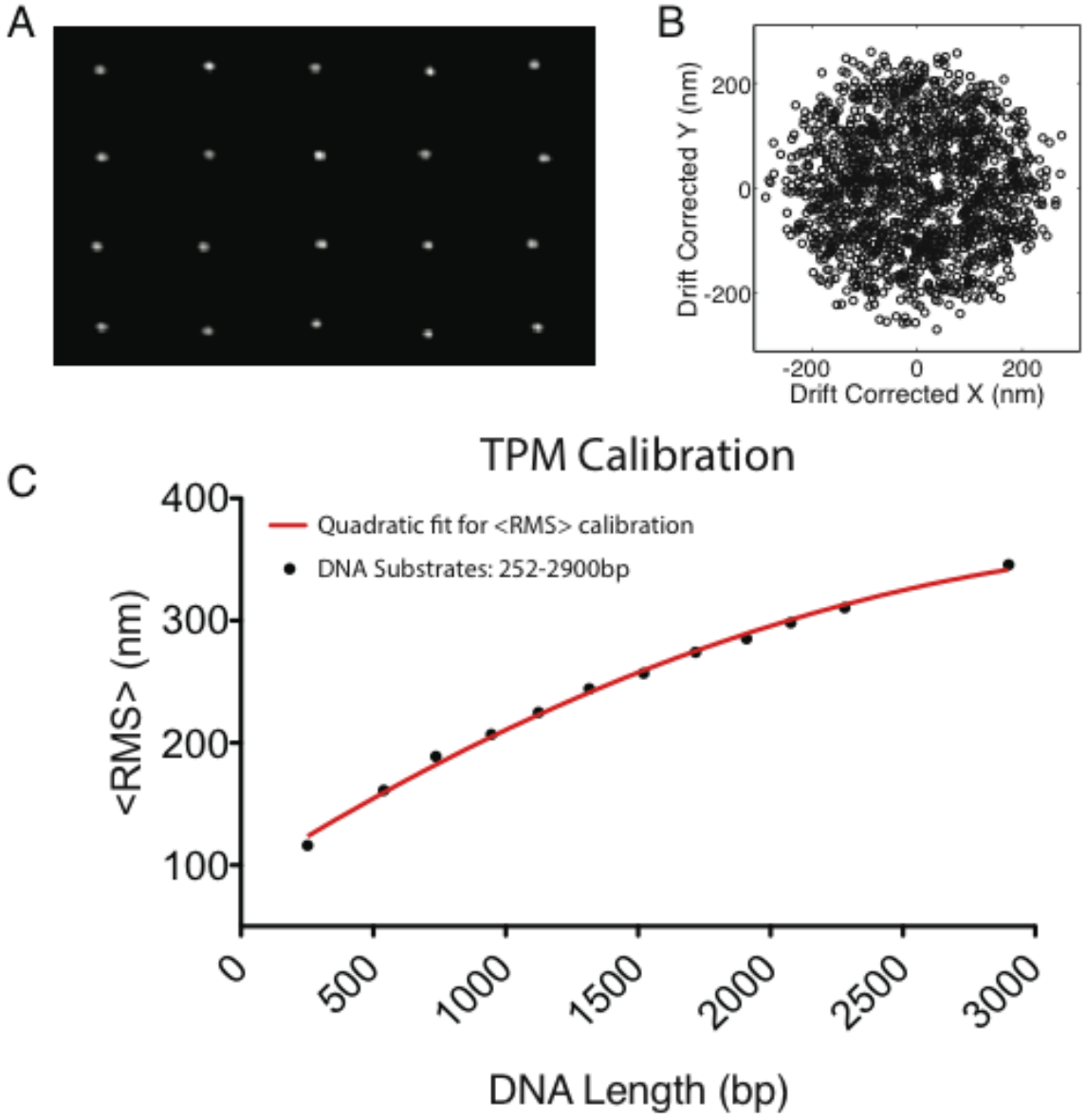

Figure 3.2: Tethered particle motion (TPM) acquisition and analysis. (A) After we assemble DNA-tethered beads we select beads in a field of view (white boxes) and keep those with a single DNA molecule coupled to them. (B) We track the Brownian motion of the DNA tethered beads generating a Brownian radius (black circles). (C) Using the Brownian distribution we extract the height or <RMS> of the DNA tethered beads as a function of length in bps (black dots). We then perform a quadratic fit to the data of the form y = Ax2 + Bx + C to extract a calibration equation (red line) where A = (-1.77± 0.3) x 10-5, B = 0.14 ± 0.01 and C = 89.9 ± 4.6. The error on A, B and C were derived from the fit to the <RMS> as a function of DNA length.

.

The addition of purified RAG1/2c to tethered DNA harboring only a single 12RSS or 23RSS binding site resulted in periods of measurable reduction in the RMS motion of the tethered beads (see example trace in Figure 3.1B). These RMS measurements are reported in terms of effective tether length (in bp) corresponding to the tether length that would produce the observed RMS motion in the absence of protein; the calibration curve for this conversion is shown in Figure 3.2C. As such, we interpret these periods of reduced RMS motion to be binding events that change the effective tether length by 64bp ± 5bp for a single 12RSS binding site and 57bp ± 5bp for a single 23RSS binding site (Figure 3.4E,F,J). Specifically, we assume that the binding of these proteins bends the DNA, changing its effective length. The lesser reduction in length for the 23RSS may be due to a reduced bend angle previously observed [18]. In Figure 3.3A this reduction in DNA tether length is visible in a histogram of measured tether length for each RAG1/2c concentration for both a single 12RSS or a single 23RSS (corresponding example bead trajectories are shown in Figure 3.4A,B). As can be seen in Figure 3.3A, at low RAG1/2c concentrations the tethers were typically measured at full length (539bp) indicating the tethers were not bound by protein. However, as the concentration of RAG1/2c was increased, a shorter bound state (at 475bp ± 5bp for the 12RSS and 482bp ± 5bp for the 23RSS) dominated the histogram. With both the 12 RSS and 23 RSS present, additional shortening of the tether beyond that already present with a single RSS present could result either from formation of the paired complex or from the presence of two separate bound complexes, each of which contributes to the overall shortening separately. We explore the case with two RSSs in the presence of RAG1/2c alone in the SI (Figure 3.4(G-J)).

Making use of this method to measure the binding of RAG1/2c to its RSS target, we compiled single-molecule trajectories to determine the probability of being bound (shortened tether) as a function of RAG1/2c concentration. Our goal in doing so is to determine the equilibrium dissociation constant (Kd) for the 12RSS and the 23RSS. Figure 3.3A already gives us an intuitive feel for Kd by estimating the concentration at which the two peaks are of equal height (and thus the concentration where either of these two states is equally likely). More formally, we employed a statistical mechanical model of binding (Figure 3.3B) [19, 20] to extract the relevant equilibrium dissociation constants (Kd) taking into account the binding histograms over the entire concentration range. In this case, for the 12 RSS site

we found Kd = 13.9nM ± 4.7nM, within error of a previously reported value of 14.5nM ± 2.4nM [21] and less than the only other measured value of Kd = 25.0nM ± 5.0nM [22]. We found the 23RSS site binds RAG1/2c considerably more weakly, with a Kd = 44.4nM ± 6.5nM (Figure 3.3C). To the best of our knowledge the dissociation constant for the 23RSS has not been previously reported. The error bars on this measurement represent the 95% confidence interval of a least-squares fit to the data in Figure 3.3C.

3.3.2 Binding kinetics reveal mean dwell time of a RAG-RSS complex

To better understand the kinetics of the binding of RAG1/2c to the 12RSS or 23RSS, we measured the average dwell time of the tether in both the bound and unbound state as a function of RAG concentration. Analyzing the same trajectories used to obtain the dissociation constant (where mean occupation probability is analyzed) we instead segmented each individual trace into bound and unbound states using a thresholding method, to visualize the kinetics of binding/unbinding events of RAG1/2c to an RSS target site. An example of a segmented trace is shown in Figure 3.1B, where the red line is the RMS trajectory of the bead and the black dashed line is the resulting segmentation between long and short states of the DNA tether.

It was of particular interest to establish the length of time that a bound molecule of RAG1/2c stays bound on a DNA molecule before dissociating. We therefore examined the mean dwell time of a RAG1/2c complex in the bound and unbound state as a function of concentration. In Figure 3.3D the mean dwell time in the bound (solid circles) and unbound (open squares) state is shown for a concentration range between 1nM and 50nM for RAG1/2c. There are several interesting features visible in this data. First, the dwell time in the bound state is roughly independent of concentration. As one might expect because of the differences in affinity, the time RAG1/2c spends bound at its target was most strongly influenced by the identity of the target to which it binds with a 430s ± 120s dwell time when bound to 12RSS (black solid points) and a dwell time of 230s ± 80s to 23RSS (red solid points). The time spent in the unbound state was a strong function of RAG1/2c concentration and was not influenced by the identity of the target binding site itself (Figure 3.3D). The full dwell time distributions for both the bound and unbound states are shown for 12RSS and 23RSS in Figure 3.6A.

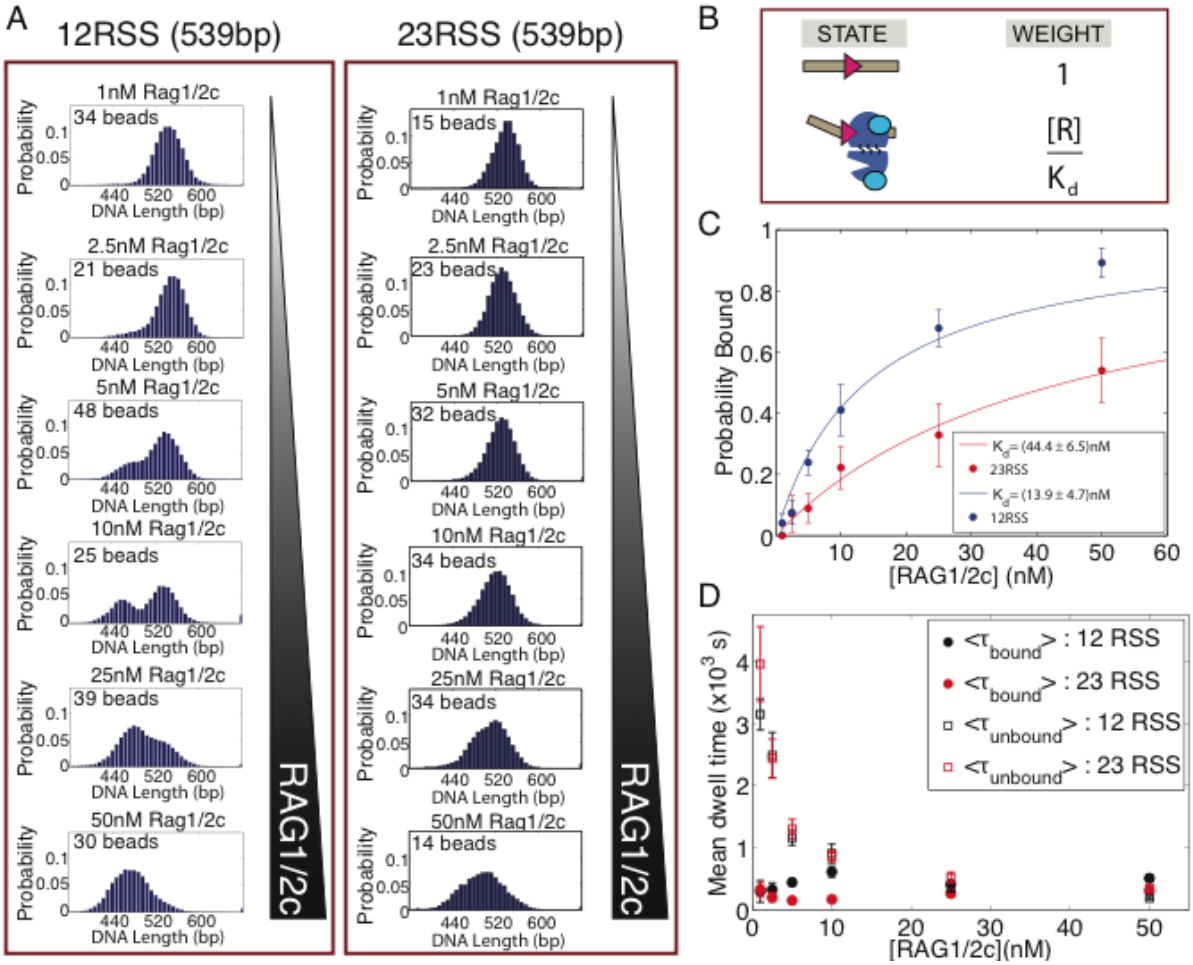

Figure 3.3: Detecting and determining the binding properties of a single RAG-RSS complex. (A) We tethered single DNA molecules containing a 12RSS or a 23RSS site, and titrated RAG1/2c from 1-50nM, and observed concentration-dependent shortening in the presence of RSSs. The number of bead trajectories in each sample is inset in each histogram. (B) A statistical mechanical model to extract the dissociation constant from the distributions shown in (A) (R = concentration of RAG1/2c, Kd = dissociation constant of RSS). (C) We fit two Gaussians to each distribution (1-50nM) and plotted the probability of being shortened as a function of (RAG1/2c). We then fit the data using the statistical mechanical model shown in (B). (D) Mean dwell times of individual RAG1/2c-RSS complexes (12RSS: black, 23RSS: red) in both the on (or bound) state (closed circles) or off (unbound) state (open squares). Binding events occurred with similar frequency independent of binding site identity, however the time bound to 12RSS was nearly twice that of 23RSS. Error bars represent standard errors calculated from bootstrap resampling.

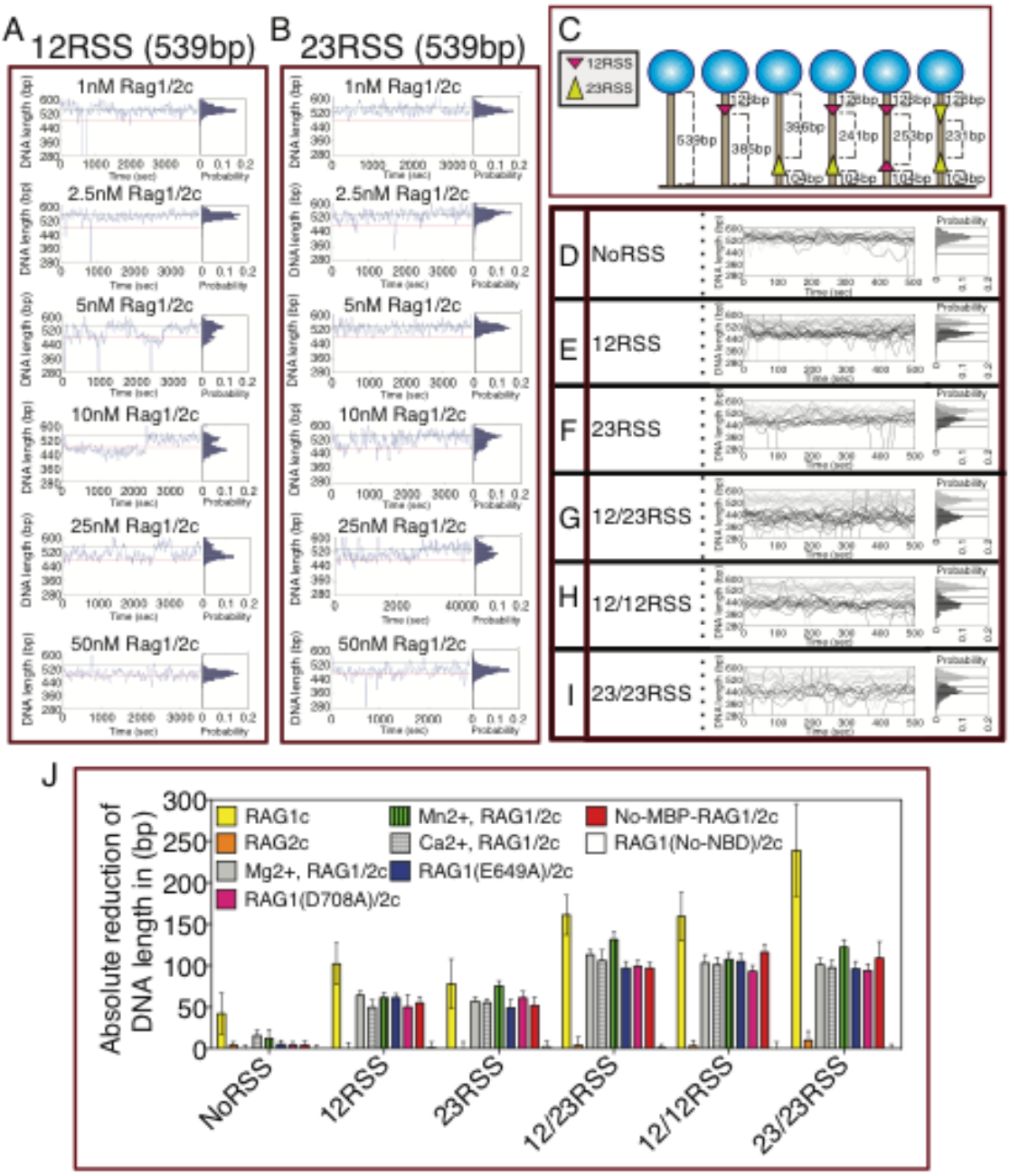

Figure 3.4: Dynamic RAG trajectories, individual RAGs and RAG mutants shortening one or two RSSs on a single DNA molecule. (A-B) RAG trajectories switching between the bound and unbound state (blue). (C) A schematic of DNAs with NoRSS, 12RSS (maroon), 23RSS (yellow), 12/23RSS combination, 12/12RSS combination and 23/23RSS combination. (D-I) Trajectories with no protein (gray) and with WT-RAG1/2c at saturation (50nM) in the presence of No, 12, 23, 12/23, 12/12 and 23/23RSSs (black trajectories). (J) 50nM RAG1c (yellow), RAG2c (orange), Mg2+ -RAG1/2c (gray bar), Ca2+ -RAG1/2c (yellow-green black-stripped), Mn2+ -RAG1/2c (white black hatched-stripped) , RAG1(E649A)/2c (blue), RAG1(D708A)/2c (maroon), No-MBP-RAG1/2c (red) and RAG1(No-NBD)/2c (white) were added.

Assuming that binding is limited by diffusion, the theoretical prediction for time to reach a target scales as the inverse of the concentration, τoff =1/(αc) where c is the concentration and α is a constant of proportionality related to the diffusivity of the particle and size of the target. In this case we found α =(2.5±0.5)x10-4 (s nM)-1. Interestingly, this number is roughly 3 orders of magnitude lower than one might expect for simple diffusion of a molecule with a hydrodynamic radius of a few nanometers to a target of similar size; while simple diffusive theory predicts these binding events should be observed at picomolar concentrations, instead we find that they occur at nanomolar concentrations. However, working concentrations at the nM scale are typical in all studies of V(D)J recombination and not unique to our assay and probably reflects the complicated nature of the RAG1/2c complex. It is possible this discrepancy is partially related to the fact that the purified RAG1/2c is not homogeneous and certainly contains some fraction of inactive or non-heterotetrameric protein. The presence of an inactive fraction of protein would shift the measured required concentrations for binding in a direction consistent with the observed trends. A possible second reason for this discrepancy is that this simple model assumes the DNA is a perfect absorber, with no constraints on molecular orientation, while in reality the binding process is probably more stringent. See the SI for further discussion.

3.3.3 HMGB1 alters the geometry of RAG-RSS complexes

In the process of V(D)J recombination, the RAG proteins do not act in isolation. For the purposes of probing the dynamics of hairpin formation, we must account for the role of HMGB1 [23]. To that end, we wished to build upon our analysis by investigating how HMGB1 altered RAG1/2c-RSS induced binding. However, this provided a challenge in that the presence of HMGB1 altered the measured DNA length even in the absence of RAG1/2c or RSS binding sites. Depending on the concentration of HMGB1 and the length of the substrate used, we observed up to 50% compaction of the DNA substrate (the <RMS> of the bead was reduced to that expected from a tether of half the actual length). Figure 3.7C shows this effect in our experiments. In the face of these condensing effects, it was imperative to verify that we could still use the TPM assay described above to measure additional RAG-RSS dependent shortening. Figure 3.5D shows a histogram of bead position over 1h in the presence of 25nM HMGB1 and varying concentrations of RAG1/2c (between

1 and 50nM). A shift in effective tether length is observed as RAG1/2c is titrated in the presence of HMGB1; at low concentration the beads exhibit a long state associated with an unbound tether, and as the concentration is increased a shorter state, associated with RAG1/2c binding to available RSS binding sites, begins to dominate. Here we add 50nM RAG1/2c both with (blue bars) and without (grey bars) 25nM HMGB1 (red bars) (Figure 3.5B), to the DNA substrates shown in Figure 3.5A, and find the mean length of a tether by fitting a Gaussian to the appropriate position histogram (like those in Figure 3.5D). Thus we see the effect of RAG1/2c binding on the observed tether length.

We exploit this observable difference in the reduction in DNA tether length to determine the dissociation constant for the 12 and 23RSS in the presence of RAG1/2c and HMGB1 using the methods described above and used in Figure 3.3B,C. One crucial caveat in the approach here is that the bound state under these conditions was very stable and in most trajectories we no longer see robust switching between the states making a measurement of the kinetic parameters impossible. However, assuming the measured state of a given tether is a representation of the overall bulk probability of being found in either of the two states, we can still calculate Kd using the thermodynamic model introduced earlier. We find that for RAG1/2c plus HMGB1 the Kds for 12RSS and 23RSS are both less than RAG1/2c alone: Kd = 4.1nM $\pm$ 1.6nM and Kd = 8.8nM $\pm$ 4.0nM for the 12RSS and 23RSS, respectively (Figure 3.5C). The presence of HMGB1 greatly enhances the binding of RAG1/2c to both the 12 and 23RSS; while it is known that HMGB1 alters the RAG1/2c affinity for the 23RSS such that it becomes comparable to that for the 12RSS by increasing RAG1/2c-23RSS-dependent DNA compaction [6, 24, 25], it was unclear if HMGB1 also enhanced RAG1/2c-12RSS DNA compaction as well (Figure 3.5B, Figure 3.7B). We also find the Kds for the 12RSS and 23RSS in the presence of RAG1/2c and HMGB1 are within error to dissociation constants previously derived from the turnover number (kcat) [26].

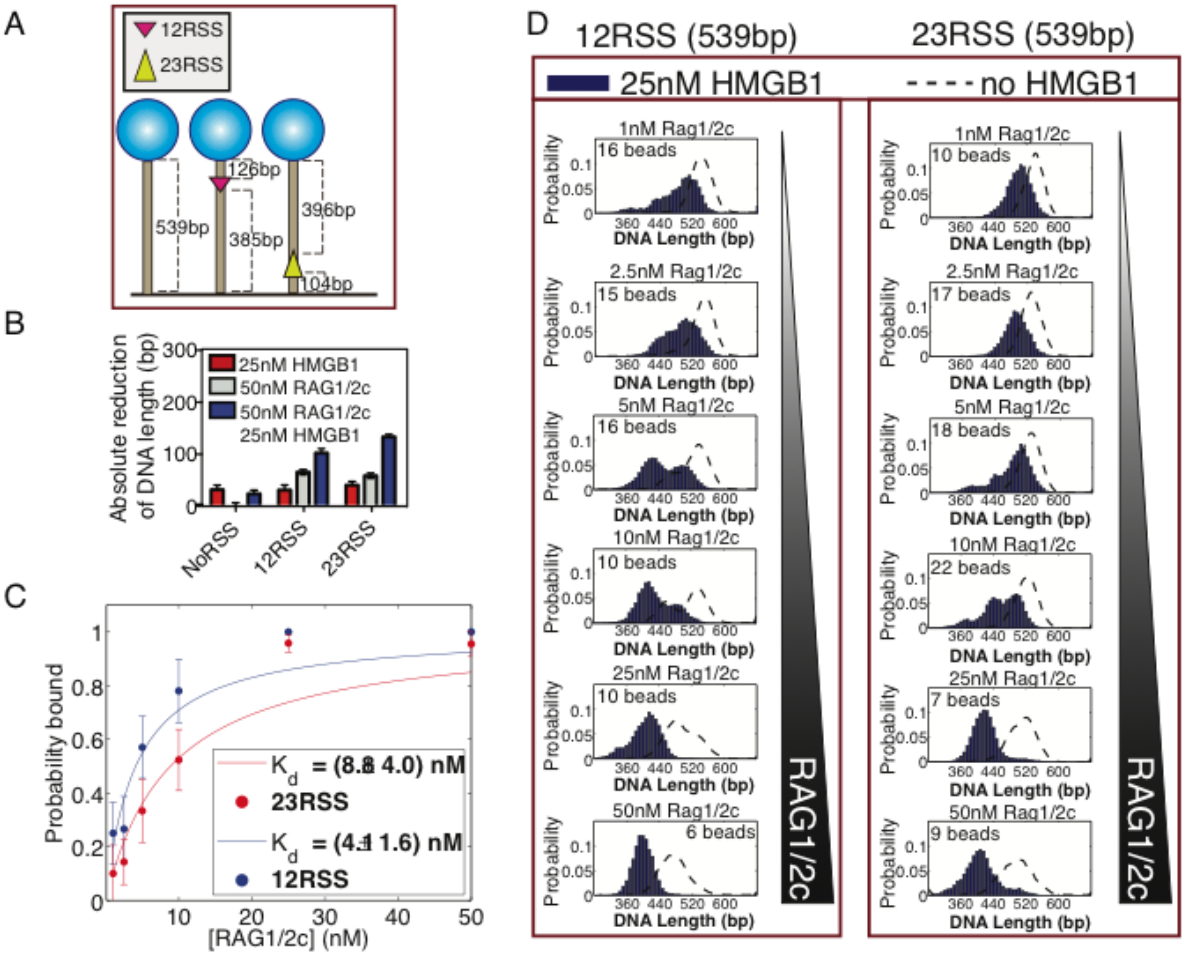

Figure 3.5: Detecting and determining the binding properties of RAG-RSS complex in the presence of HMGB1. (A) A schematic of DNAs with NoRSS, 12RSS (maroon) and 23RSS (yellow). (B) 50nM RAG1/2c with (blue) or without (gray) 25nM HMGB1 (red) was added to the DNA substrates shown in (A). (C) We fit two Gaussians to each distribution (1-50nM RAG1/2c and 25nM HMGB1, as seen in (D) and plotted the probability of being shortened as a function of both RAG1/2c and HMGB1. (D) To calculate Kd for 12RSS and 23RSS in the presence of 25nM HMGB1, we tethered single DNA molecules containing a 12RSS or a 23RSS site, and titrated RAG1/2c from 1-50nM in the presence of 25nM HMGB1, and observed concentration-dependent shortening in the presence of RSSs. The number of bead trajectories in each sample is inset in each histogram, and the dotted black line is the effect of RAG1/2c alone.

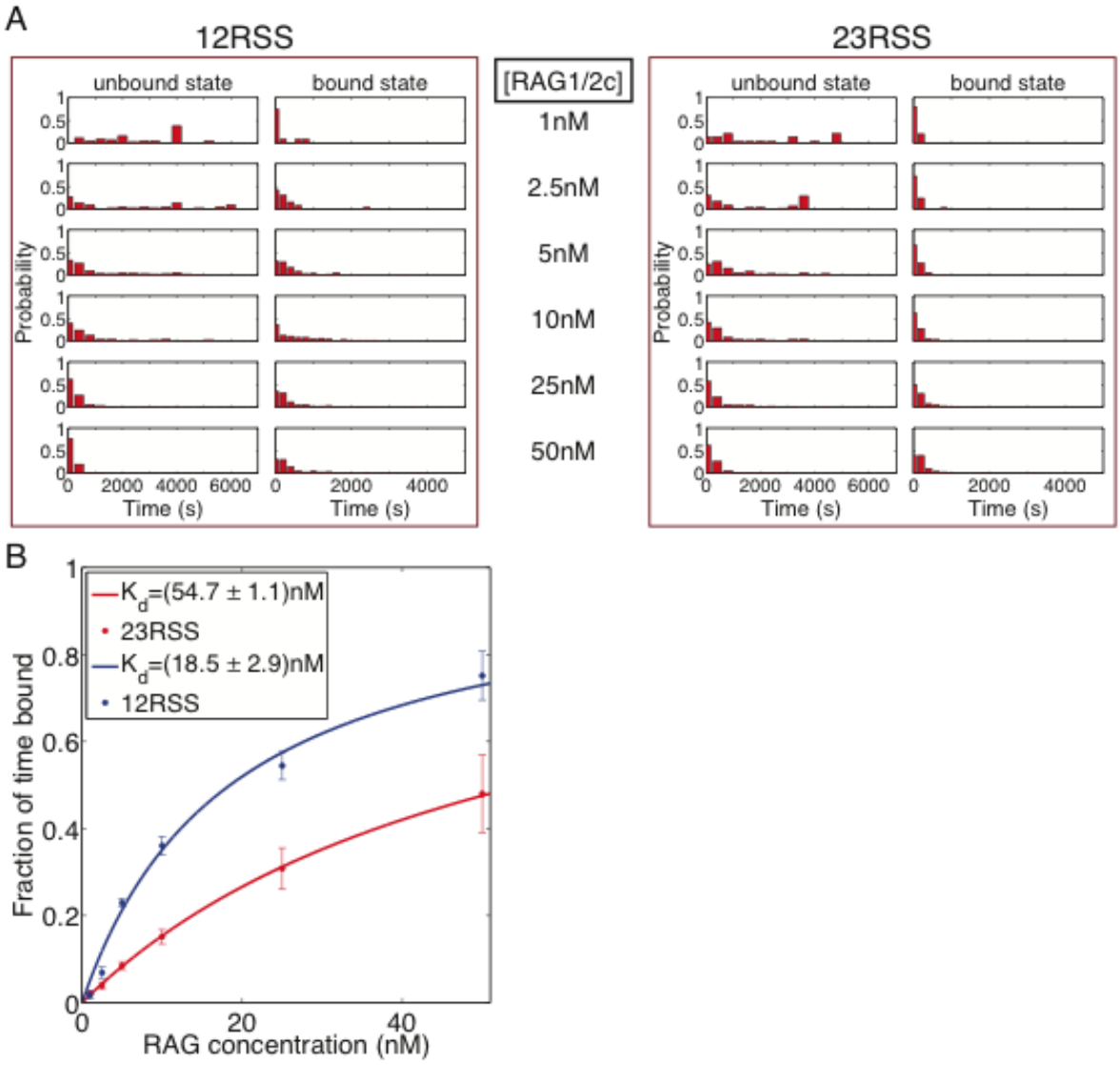

Figure 3.6: Using fractional occupancy to determine dissociation constants. (A) Histograms of individual dwell times measured with 12RSS (left) and 23RSS (right) for RAG1/2c concentrations between 1nM and 50nM. (B) The fractional occupancy of RAG1/2c as a function of RAG concentration (12RSS = blue, 23RSS = red). This is calculated by summing all the time spent in the on state over every bead and dividing by the total time of measurement. To extract the Kd for a given binding site, the fractional occupancy curves are fit vs. the concentration of RAG1/2c, c, to a function of the form PBound = (c/ Kd)/(1+(c/ Kd)) (a Langmuir isotherm or Hill function with n=1). These fits gave Kd = 18.5nM ± 2.9nM for 12RSS and Kd = 54.7nM ± 1.1nM for 23RSS.

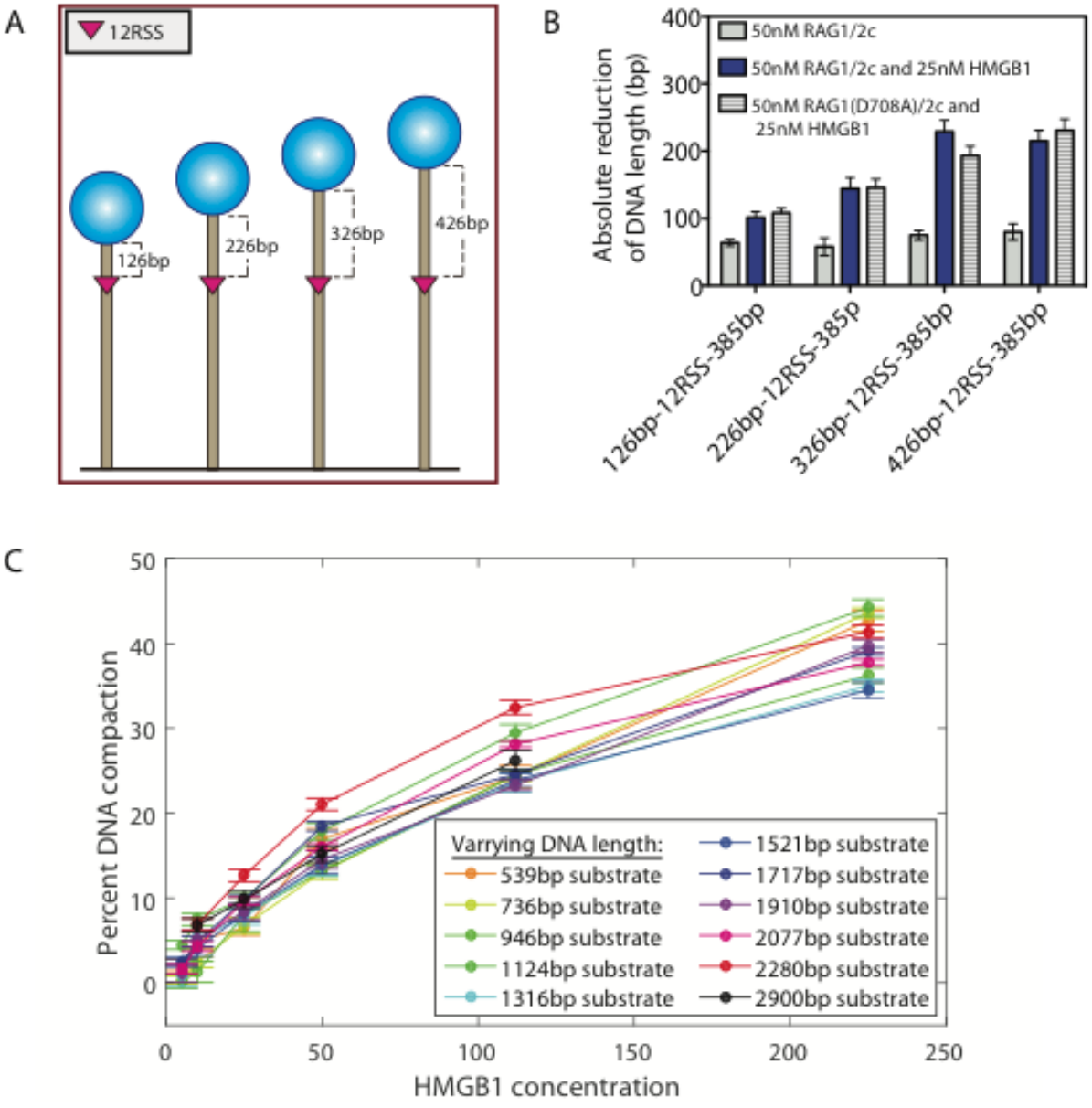

Figure 3.7: Length dependence of RAG-HMGB1-RSS complex formation. (A) A schematic of DNA substrates to calibrate the optimal DNA length flanking RSSs to observe complete RAG and HMGB1 bends. The length of DNA downstream of the 12RSS (maroon) is fixed at 385bp and we increase DNA upstream to 126bp, 226bp, 326bp and 426bp. (B) 50nM RAG1/2c (gray), 50nM RAG1/2c and 25nM HMGB1 (blue) and 50nM RAG1(D708A)/2c and 25nM HMGB1 (white stripes), was added to the DNA substrates shown in (A). (C) The presence of HMGB1 altered the measured effective length of DNA tethers in a concentration dependent way. This effect was observed in the absence of RAG1/2c and was found to depend on the overall length of DNA substrate.

3.3.4 Direct observation of paired complex formation

The paired complex is predicted to form prior to hairpin production (observed as bead loss in our experiments). To determine if we could observe the paired complex prior to bead release, we tracked beads attached to a 12-23RSS substrate with intersignal spacing of either 1200bp or 1800bp in the presence of 5nM RAG1/2c and 80nM HMGB1 for 1h (Figure 3.8A). Importantly, a paired complex that forms on either of these two substrates (which have identical overall lengths) and loops out the intersignal DNA should display distinctly different DNA lengths (1360bp for the 1200bp intersignal substrate and 880bp for the 1800bp intersignal substrate), allowing us to be certain that an observed shortening event is coming from a paired complex connecting the RSS sites, as shown schematically in Figure 3.8A. Furthermore, as a control we performed the same experiment on the NoRSS and 12RSS substrate where we would not expect to see paired complex formation. Indeed, on these control substrates we did not see any instances of the paired complex state.

An example of the observed trajectories is shown in Figure 3.8B where the paired complex state can be seen through the existence of a shortened state in several trajectories both for 1200bp intersignal spacing (top traces) and 1800bp intersignal spacing (see Figure 3.9A-D for trajectories). These states end in one of two fates: 1) a cutting event where the DNA is cleaved and the bead immediately leaves the field of view (i.e. the bead does not transition back to the unlooped length before release) or 2) a return to the previous unlooped state and the tether is not cleaved (the bead returns to the unlooped length although it may be cleaved later either passively or by another complex formation). In the insets of Figure 3.8C, we show the histogram of median effective DNA tether length (in bp) for these states (black; cleaved complexes, red; no cleavage) for both substrates where it is clearly visible that the DNA tether length of the observed short state depends on the length of intersignal spacing between RSSs. Furthermore, the effective DNA tether length of the paired complex state agrees with the predicted length assuming the substrate had shortened by 1200 or 1800bp, respectively, and the remaining DNA is compacted 20% by HMGB1 (as expected from figure 3.7C). A histogram of the dwell time in the paired complex for states ending in cleavage (black) and no cleavage (red) is shown in Figure 3.8C. The dwell times are relatively short, on the order of hundreds of seconds (although

we did rarely observe paired dwell times of thousands of seconds)and it can be seen that three quarters of all observed paired complex states are aborted without cleavage. One interesting point that arises in these measurements is demonstrated in Figure 3.8D. Here we show a histogram of the probability of bead loss for the 12-23RSS (1200bp and 1800bp), 23RSS only and NoRSS substrate; that is, the probability that any one bead will float away during the course of the experiment. There are two clear mechanisms for bead loss in our experiments; the DNA cleavage event that follows from paired complex formation or from typical but unpleasant passive reactions that have nothing to do with the nature of the DNA substrate or the presence of proteins. For the NoRSS and 12RSS the rate of bead loss are roughly the same and we attribute this to this passive form of bead loss. However, for the 12-23RSS substrates there is a population of bead loss events that are immediately proceeded by a measurable paired complex, this fraction is colored black in the histogram bars. Furthermore, the substrates with both RSS sites (1200 bp and 1800 bp substrates) seemed to lose more beads than the control substrates (12RSS only and NoRSS) without a paired complex signature; i.e. the blue bar in Figure 3.8D is bigger with both RSS sites present. We do not yet know the reason for this but we believe it is likely due to 1) paired complex formation and excising the looped DNA on timescales faster than our resolution or 2) a pathway for cutting that does not require the two RSS sites to synapse.

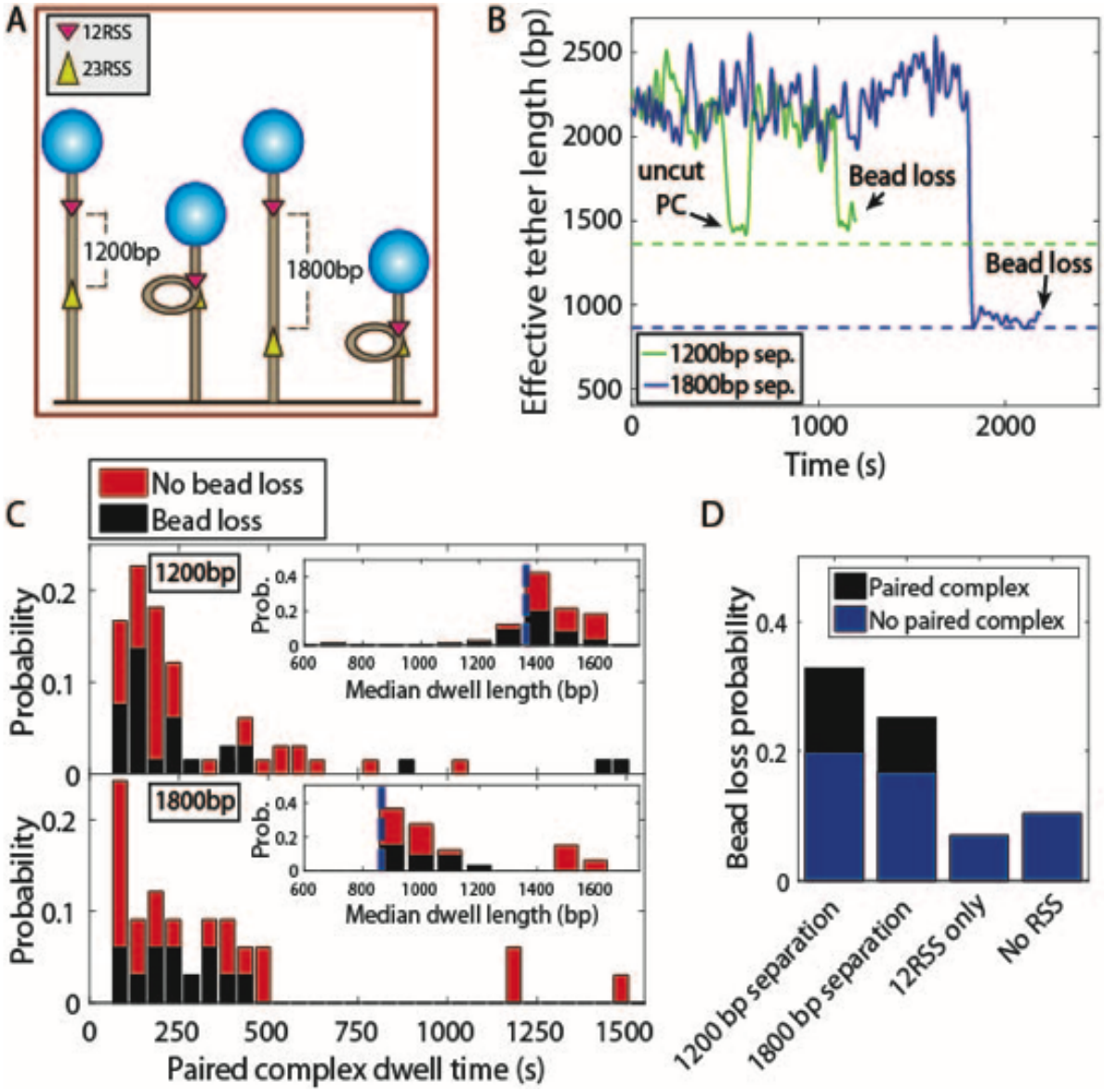

Figure 3.8: Capturing the dynamics of paired complex formation. (A) Schematic of the DNA substrates used to investigate the dynamics of paired-complex formation. The use of different lengths results in different characteristic TPM signatures upon paired complex formation. (B) TPM trajectories showing the dynamics of paired complex formation. The green trajectory is from the substrate where 12RSS and 23RSS are separated by 1200bp and the blue trajectory is from the substrate with 1800bp separation. The dashed, colored lines are the expected tether lengths in bp for paired complex formation, where the DNA between the RSS sites is looped out. Each sample trajectory terminates in a state corresponding to the correct length for a paired complex before a cleavage event resulting in bead loss. For the 1200bp substrate, a complex that does not result in cleavage is seen earlier in the trajectory. (C) Lifetimes of the paired complex for the case in which the complex forms reversibly (red) and in the case when the paired complex is the terminal state prior to bead release (black). (D) Histogram of bead loss events showing the percentage of bead loss, the number of beads lost passively and those lost as a result of paired complex formation for the NoRSS, 12RSS, 12-23RSS-1200bp or 12-23RSS-1800bp substrates.

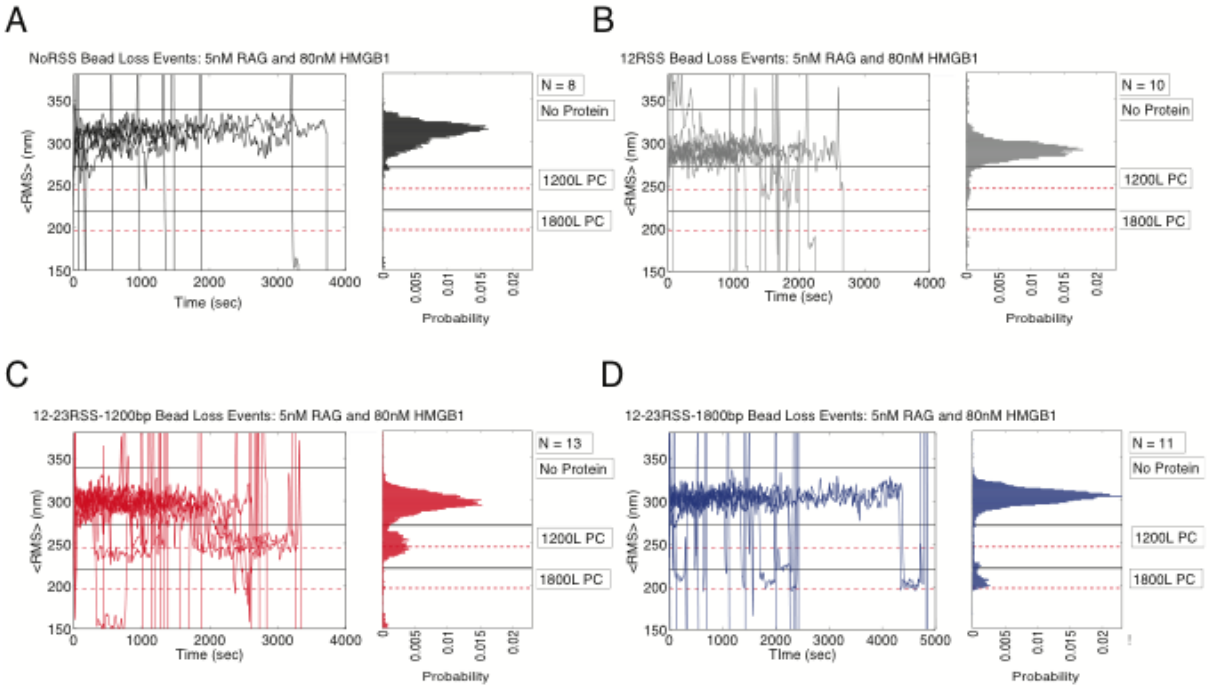

Figure 3.9: Composite trajectories and normalized probability distribution for NoRSS, 12RSS, 12-23RSS-1200bp and 12-23RSS-1800bp DNA substrates. (A,B) Trajectories and distributions for the NoRSS (black) and 12RSS (grey) beads resulting in bead loss. (C,D) Trajectories and distributions for the 12-23RSS-1200bp (red) and 12-23RSS-1800bp (blue) beads showing the paired complex and resulting in bead loss (Top black solid line: no protein state, two lower black lines: predicted paired complex without HMGB1 for 1200bp and 1800bp, and two lower red dotted lines: predicted paired complex with 80nM HMGB1 for 1200bp and 1800bp).

3.3.5 The dynamics of 12/23 rule-regulated bead release as a function of RAG1/2c and HMGB1

As discussed above, the role of the RAG recombinase is to initiate the process of V(D)J recombination by formation of the paired complex followed by cleavage (hairpin formation) at the border of each RSS. In our TPM experiment, when this process is carried out it will cleave both strands of the DNA substrate allowing the bead to drift from the field of view (shown schematically in Figure 3.10A). To quantify the role of concentration of the molecular players on the rate of this cleavage, we next used bead loss as a marker to enable time resolved, single-molecule measurements of the statistics of DNA cleavage.

The rate of bead release was quantified by observing a population of beads, and determining how many are left in the field of view as a function of time. First, to measure 12/23RSS-dependent bead release, we examined 70-80 DNA tethered beads with a 12/23RSS binding site substrate and compared it to bead release with only one or the other RSS. To control for natural, passive bead release, e.g. due to dissociation of DNA-surface interactions that occurs slowly over time, we also performed a control measurement on DNA tethered beads lacking any RSS binding site, but with the RAG1/2c and HMGB1 proteins present in the same concentrations. As seen in Figure 3.10C, significant bead release only occurs in the case where both a 12RSS and 23RSS are present simultaneously, confirming 12/23-rule regulated cleavage in our single-molecule assay. No other combination of RSS sites on our DNA substrate (12RSS, 23RSS, 12/12RSS and 23/23RSS) produced significant bead release (Figure 3.10C).

How do all of these results depend upon the quantity and identity of the relevant proteins? The dependence of these results on HMGB1 concentration is shown in Figure 3.10D. In particular, we observed an increase in the frequency of RAG1/2c-mediated 12/23RSS bead release with increasing concentrations of HMGB1 between 5nM and 500nM. We also found that hairpin production in a bulk assay (Figure 3.11A) and bead release in our single-molecule assay (Figure 3.11B,C), required catalytically active RAG1/2c, consistent with previous bulk analyses [1].

We were also interested to see if there was additional bead release activity in the presence of more than one 12RSS as a first in vitro step towards the cellular situation in

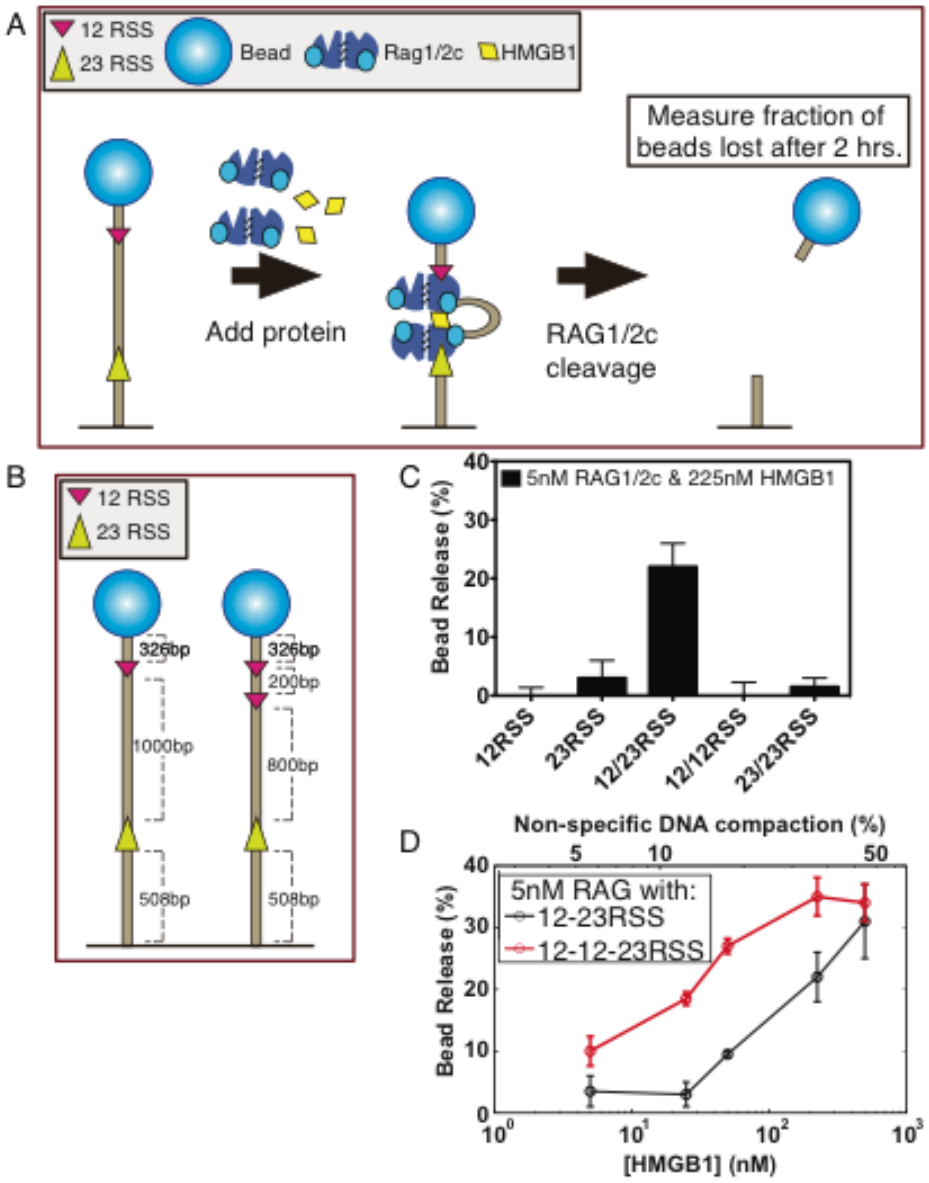

Figure 3.10: 12/23 rule regulated bead release as a function of HMGB1 concentration. (A) A schematic of the assay to measure 12/23 rule regulated hairpin production by bead release in the TPM assay. (B) Schematic of 12/23RSS and 12-12/23RSS substrates used in (C). (C) Bead release as a function of HMGB1 concentration in the presence of 5nM RAG1/2c for 12/23RSS and 12-12/23 DNA substrates. The upper axis corresponds to the percent non-specific DNA compaction at that concentration of HMGB1. The percentage bead loss for the 12/23RSS was determined using the equation $\triangle$NBR/Ntot = $\triangle$Ntot/Ntot - $\triangle$NPBR/Ntot where NBR is number of beads released, NPBR is the number of beads released passively, and Ntot is the total number of beads lost. The percent non-specific DNA compaction is derived from the change in <RMS> at different HMGB1 concentrations on the NoRSS DNA substrate (see Figure 3.7C). (D) Significant bead release was only detected in the presence of a 12/23RSS substrate (experiments used 5nM RAG1/2c and 50nM HMGB1).

which there are multiple RSSs. The red points in Figure 3,10D show the measured bead release when a second 12RSS site is added to the DNA substrate as depicted in Figure 3.10B. In this case, where there are now two configurations for 12/23RSS hairpin production, bead release is increased although it appears to saturate at roughly 30%, similar to that with the standard 12/23RSS substrate. Figure 3.10C shows the fraction of bead loss in the presence of 5nM RAG1/2c and 225nM HMGB1 after two hours on a number of different DNA substrates.

One very interesting parameter that governs the frequency of bead release is the distance between the two RSSs. Bead release activity can be ablated by decreasing inter-RSS separation from 137bp to 73bp (Figure 3.11(D, E)), which would be predicted to interfere with RSS synapsis [27] and has been shown to reduce cleavage in biochemical assays [28]. Notably, the block to bead release imposed by the 73bp inter-RSS distance is overcome when Mn2+ is used in place of Mg2+ in the reaction buffer (Figure 3.11F). Mn2+, unlike Mg2+, allows hairpin formation in the absence of paired complex formation [29]. Overall, these results regarding the combination of RSSs, reaction conditions, and intersignal distance required for bead release are in close agreement with results from bulk studies of RAG-mediated DNA cleavage and provide additional confidence that bead release is the result of hairpin formation in the paired complex as seen in typical bulk experiments.

Figure 3.12A-C shows the time evolution of bead release over two hours for three different concentrations of RAG1/2c (1nM, 5nM and 25nM) and a range of HMGB1 concentrations between 5nM and 500nM. First, we see that nearly all release activity occurred in the first hour with at least 5nM RAG1/2c, a concentration similar to Kd for both the 12RSS and 23RSS in the presence of HMGB1 (Figure 3.5D). Second, the resulting bead release dynamics seemed to fall broadly into one of two categories; bead release occurs quickly over the first hour before saturating around 30% (for instance, at 5nM RAG1/2c with 225 or 500nM HMGB1 or 25nM RAG1/2c with 50, 225 or 500nM) or bead release is not strongly observed and stalls below 10% (as is the case for lowest two concentrations of HMGB1, regardless of RAG concentration, and for the lowest RAG concentration).

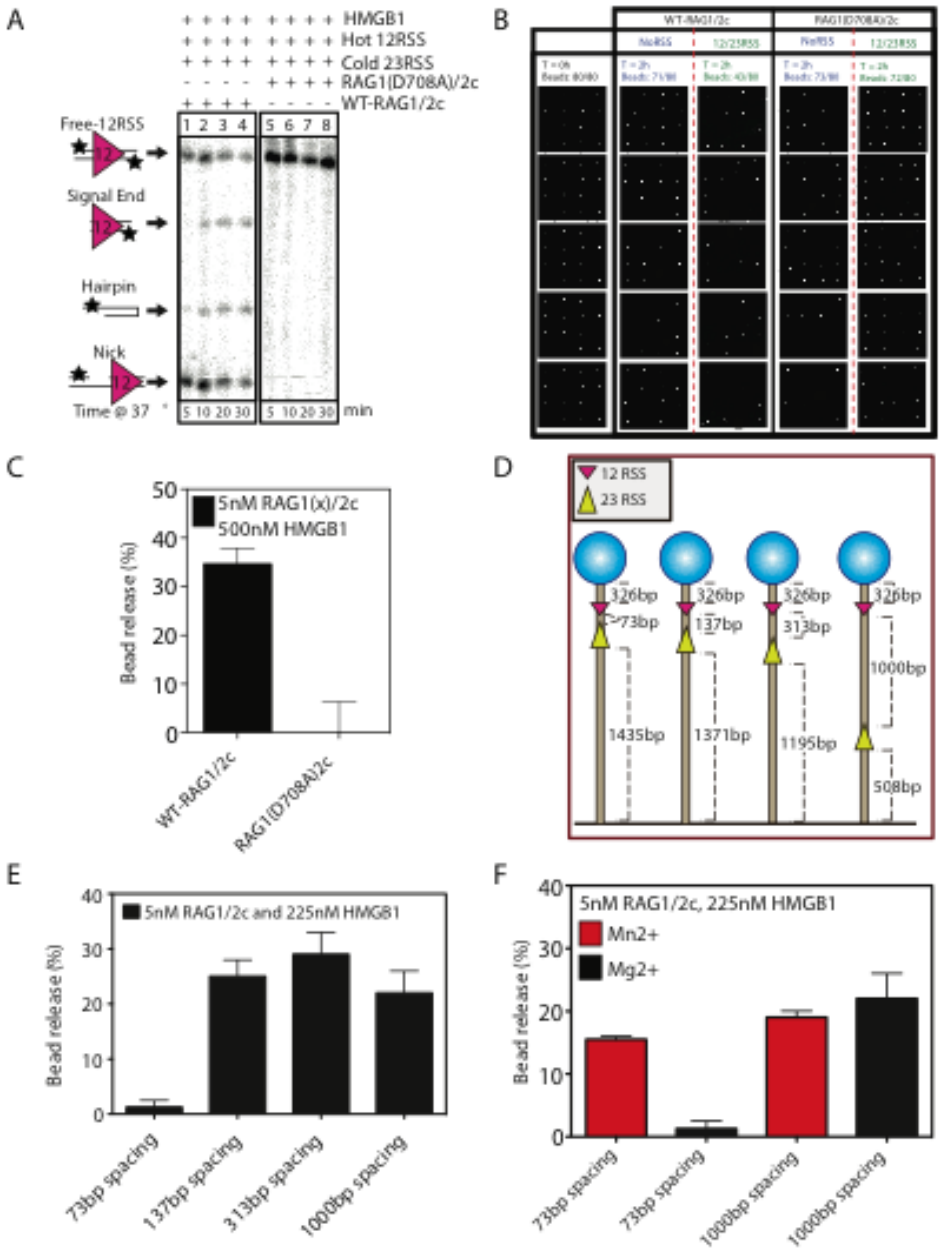

Figure 3.11: RAG and HMGB1 ensemble and single-molecule in vitro cleavage assays. (A) 50nM RAG1/2c (lanes 1-4) or RAG1(D708A)/2c (lanes 5-8) and 225nM HMGB1 was added to 1nM radiolabelled 12RSS oligo (50bp), then 4nM 23RSS was added in trans. The reaction was incubated at 37C and samples were removed at 5, 10, 20 and 30 min. The samples were fractionated on a sequencing gel, to detect nicks and 12/23 rule regulated hairpin production. (B) In the single-molecule assay for hairpin production we added 5nM RAG1/2c or RAG1(D708A)/2c and 500nM HMGB1 to NoRSS (blue) or 12/23RSS (red). We determined % bead release using the equation $\triangle$NBR/Ntot = $\triangle$Ntot/Ntot - $\triangle$NPBR/Ntot , where NBR is number of beads released, NPBR is the number of beads released passively, and Ntot is the total number of beads. (C) A schematic of DNA substrates with 12 and 23RSS spacings of 73, 137, 313 and 1000bp. (D) We added 5nM RAG1/2c and 225nM to the four DNA substrates shown in (C) and assayed for bead release. (E) We plot the % bead release for the experiment shown in (B) for RAG1/2c or RAG1(D708A)/2c (black bars). (F) Lastly, we add 5nM RAG1/2c and 225nM HMGB1 in the presence of 12/23RSS substrates with 73bp or 1000bp spacing in the presence of Mn2+ (red) and Mg2+ (black).

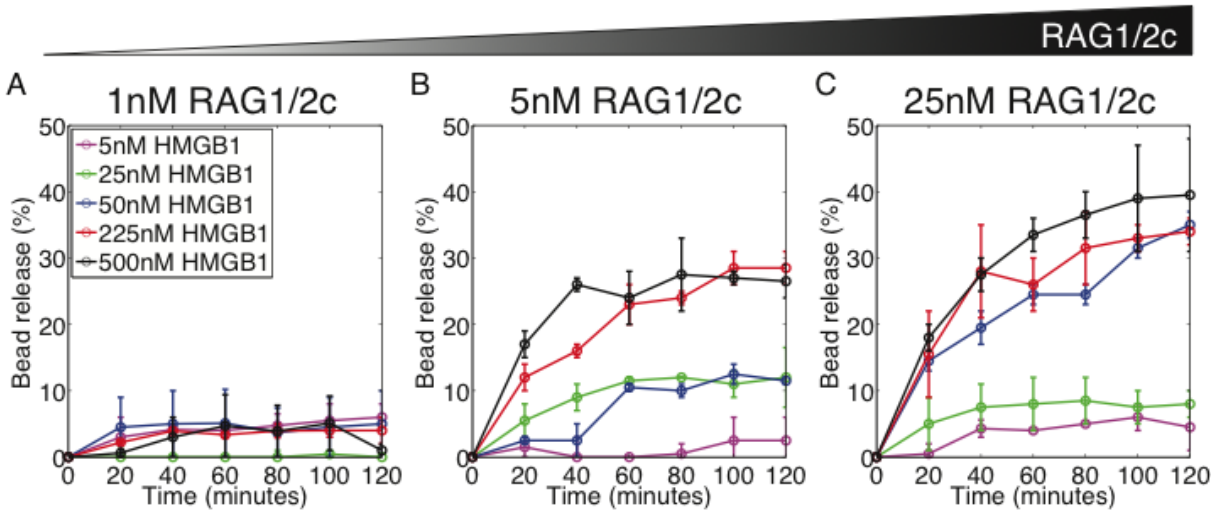

Figure 3.12: Dynamics of 12/23 rule-regulated bead release at 1nM, 5nM and 25nM RAG1/2c concentration. Hairpin formation as measured by bead release as a function of time measured over two hours. (A) At 1nM RAG1/2c, hairpin formation did not occur appreciably over the two-hour window we observed. (B) For 5nM RAG1/2c, the highest concentrations of HMGB1 (225nM and 500nM) produced hairpins quickly over the first hour but hairpin formation slowed at around 25% formation. The intermediate concentrations now displayed small but measurable bead-loss. (C) With 25nM RAG1/2c, the dynamics of two highest HMGB1 concentrations were not changed, however 50nM HMGB1 now displayed very similar dynamics to these high concentration curves. The lowest concentrations (5nM and 25nM) remained relatively unaffected.

3.4 Discussion

V(D)J recombination is one of the most intriguing examples of the striking dynamics of genomes. The specific mechanistic workings of such genome dynamics still entail many unanswered questions. To address some of these unanswered questions at the single-molecule level, we aimed to develop a single-molecule assay that would make it possible to watch the various steps in the process in real time. Here, we have shown that we can capture individual RAG-mediated cleavage reactions from start to finish in real-time, revealing important parameters of DNA binding and cleavage by RAG and HMGB1. We observe RSS-binding by RAG in the absence of HMGB1, which agrees with previous work showing that RAG proteins alone bend DNA at RSSs [18]. This provided a dissociation constant for RAG-RSS. The measured Kd for the consensus 12RSS is within error of a previously reported value; we were also able to measure a Kd for the consensus 23RSS. Our single-molecule approaches for measuring Kd could be used to measure dissociation constants for non-consensus RSSs to provide a more complete understanding of the relationship between RSS sequence and RAG activity. Indeed in preliminary studies we have already seen that concentrations of RAG that saturate consensus RSSs show significantly reduced binding for non-consensus RSSs derived from the immunoglobulin heavy chain locus.

This approach has also allowed a first determination of the mean dwell time for the RAG1/2c complex bound to either a 12 or 23 RSS. We find that the RAG1/2-RSS complex is quite stable with dwell times on the order of several minutes. In particular, we found the complex stayed bound longer at 12RSS sequences (430 s) compared to 23RSS (230 s), consistent with the difference in Kd between the two sites; the measured difference in Kd between 12RSS and 23RSS is consistent with the idea that changing the binding site primarily changes the stability of the complex and not the time it takes to bind the RSS target (Kd12RSS/ Kd23RSS kbound23RSS/ kbound12RSS), where kbound is the off-rate for unbinding from the RSS.

We found that HMGB1 increases the reduction in DNA length of a RAG-RSS complex. This is consistent with the HMGB1-dependence of the large bend detected in 23RSS-RAG-HMGB1 complexes in an earlier work, although that study focused primarily on the 23RSS in the paired complex and used short oligonucleotide substrates [25]. We also used this reduction in DNA tether length to determine the dissociation constant for the 12RSS

and 23RSS in the presence of RAG1/2c and HMGB1 and found they both decrease compared to the case of RAG1/2c alone and are similar, consistent with previous observations [6, 26]. Our single molecule assay to study HMGB1-dependent proteins could be extended to study other proteins besides RAG1/2c that are facilitated by HMGB1, for example the tumor suppressor proteins p53 and RB.

At our concentrations of RAG1/2c and HMGB1, the binding events associated with RAG1/2c in the presence (or absence) of HMGB1 were relatively frequent however the downstream product of this binding, paired complex formation, was comparatively rare; in our experiment the bead spends just 4% of the total experiment time in the paired complex state. We directly observed paired complex formation on DNA substrates with both a 12RSS and 23RSS. We do not see evidence of this complex when only a 12RSS (or no RSS) is present and we confirm that the apparent length of the paired complex state is consistent with a tether that has been shortened by the 12/23 intersignal spacing of the substrate when using both a 1200bp and an 1800bp spacing. The paired complex was relatively short lived, which is at odds with kinetic models derived from cleavage experiments with oligonucleotide RSS substrates that predict slow hairpin formation with a long lived paired complex [26]. Furthermore, the paired complex did not always result in a cleaved DNA product; 72%of paired complexes do not lead to DNA cleavage. Our preliminary conclusion from this reversibility is that, in vivo, the RAG recombinase could sample multiple RSSs adjacent to particular V or D gene segments after establishing the recombination center at the J gene segments of immunoglobulin and T-cell receptor loci, which could impact the relative abundance of a particular VDJ or VJ alleles [30]. Our assay positions us to measure the mean dwell time of every 12/23 rule-regulated V, D and J assembly reaction carried out to generate all immunoglobulins and T-cell receptors during lymphoid development.

Finally, we turned to study RAG-HMGB1-RSS mediated cleavage on the DNA by inferring successful hairpin events by observing bead release events. With this assay, we quantified the relationship between bead release and the concentrations of the molecular players RAG1/2 and HMGB1 in cleavage. We found that bead release was greatly stimulated in a substrate containing a 12/23RSS pair, required catalytically active RAG, and was strongly inhibited when the intersignal distance was shortened beyond a certain point

previously shown to inhibit RAG-mediated cleavage and V(D)J recombination.

With these tools in hand, it is now possible to perform a number of systematic studies of the V(D)J reaction process. Our results already hint that both the number of recombination signal sequences and their distances conspire to alter recombination dynamics and it will be of great interest to determine the quantitative rules governing synapse and hairpin formation.

3.5 Methods

3.5.1 Large scale purification of cMR1/cMR2 from HEK 293-E cells

RAG proteins were purified as previously described with a few modifications [14].WT and mutant RAG proteins were purified from suspension 293-E cells using CMV driven pTT5-MBP-RAG1c and RAG2c expression vectors. First we transformed pTT5-MBP-RAG1 or 2c harboring ampicillin resistence in Z-competent cells (Zymoresearch) and plated on LB-agar containing 50μg/mL carbenicillin (Sigma-Aldrich). A single colony was selected and a large scale plasmid purification was performed using the PowerPrep HP Plasmid Maxiprep kit (Origene). Next 1L of suspension 293-E cells were grown to a cell density of 1 x 10^6 cells/mL, prior to transfection in Gibco freestyle 293 expression medium (Life Technologies). One liter of 293-E cells were transfected with 450μg of pTT5-MBP-RAG1c, 550μg of pTT5-MBP-RAG2c and 1.0mL of BioT (Bioland) brought to a final volume of 35mL in 1X PBS (Note: Mix DNA and 1X PBS first). The transient transfection proceeds for 48hrs and the cells are harvested at 3,000 rpm (700g) for 15 min. Next the cell pellet is resuspended in 30mL of lysis Buffer A (10mM sodium phosphate pH 7.4, 500mM NaCl, 1mM DTT, 0.25% Tween-20, 10μg/mL aprotinin, 10μg/mL leupeptin, 10μg/mL pepstatin and 100μM phenylmethanesulfonylfluoride (PMSF)) and microfluidized on ice. The cell lysate is centrifuged at 86,000g for 40 min using the SW28 rotor (Beckman). The supernatant is then added to 2mL of amylose resin (New England Biolabs) and inverted for 1h 30 min. The amylose resin/supernatant are poured into a 15mL poly-prep gravity flow column (Bio-Rad). The amylose resin complex was washed with 20mL of Buffer A, followed by 10mL of Buffer B (10mM sodium phosphate pH 7.4, 500mM NaCl, 1mM DTT), 300μL of elution buffer (10mM sodium phosphate pH 7.4, 500mM NaCl, 1mM DTT and 10mM maltose) to

displace column void volume and finally 2mL of elution buffer. The protein was dialyzed for three hours in dialysis buffer (25mM Tris-HCl (pH: 8.0), 150mM KCl, 2mM DTT and 10% glycerol) using a 0.5 to 3mL slide-a-lyzer (Pierce), nap-frozen and stored at $-80\,^{\circ}$C.

3.5.2 Purification of HMGB1

HMGB1 was purified as previously described with slight modifications [14]. First human HMGB1 was codon-optimized for bacterial expression by DNA2.0 and cloned into pJExpress vector. pJExpress-HMGB1 was transformed into BL21(DE3)pLysS cells and plated on LB-agar with 50μg/mL kanamycin (kan). A single clone was picked for overnight liquid culture in 50mL LB-kan, the following day 30mL of the overnight culture was diluted 100-fold in 3L of LB-kan. The cells were grown to an OD600 of 0.7 and induced by 1mM IPTG at 30 °C for 4h . The cells were harvested at 8,000rpm for 30 min and resuspended in 30mL of binding buffer (BB) (40mM Tris-HCl pH 8.0), 500mM KCl, 0.25% Tween-20 and 50mM imididazole, 10μg/mL aprotinin, 10μg/mL leupeptin, 10μg/mL pepstatin A, 100μM PMSF and 5U/mL benzonase) for every 1L of cells. The resuspended cells are passed through a microfluidizer three consecutive times to lyse them. The cell lysate is centrifuged at 46,000g for 40 min and the cleared lysate is loaded on to a 5mL Ni-NTA column. The column is washed with 40mL of BB without benzonase followed by 20mL of wash buffer (40mM Tris-HCl pH 8.0, 65mM KCl, 50mM Imidazole) and eluted with 30mL elution buffer (40mM Tris-HCl pH 8.0, 65mM KCl, 500mM Imidazole). Next we used cation (SP) and anion (Q) exchange chromatography to eliminate truncated forms of HMGB1 using buffer solution A (Soln A: 40mM Tris-HCl (pH 8.0)) and solution B (Soln B: 40mM Tris-HCl (pH 8.0) and 1M KCl). The SP-column was equilibrated in 8% Soln B or 80mM KCl, the Ni-NTA eluate was loaded onto the column and washed with 6mL of 8% Soln B and eluted in 1mL fractions with a gradient of 25mL Soln B from 8% to 65%. The fractions with HMGB1 were pooled and diluted to 80mM KCl, and loaded on a cation exchange Q-column equilibrated with 8% Soln B. The column was washed with 7mL of 8% Soln B, and eluted in 1mL fractions with a gradient of 40mL Soln B from 22% to 70%. The 1mL fractions with HMGB1 were pooled, concentrated, snap-frozen and stored at $-80\,^{\circ}$C.

3.5.3 Oligonucleotide and long DNA substrates

Consensus 12RSS (DAR39 and DAR40) and 23RSS (DAR61 and DAR62) oligonucleotide substrates were used. The sequence of the sense strand of the 12RSS-DAR39 is 5-GATCTGG CCTGTCTTACACAGTGCTACAGACTGGAACAAAAACCCTGCAG-3 and 23RSS-DAR61 is 5-GATCTGGCCTGTCTTACACAGTGGTAGTACTCCACTGTCTGGCTGTACAAAAA CCCTGCA-3. The long DNA substrates of various lengths used in bulk and single molecule assays were created using PCR and cloned into a moloney murine leukemia viral vector pOOC- (NoRSS, 12RSS, 23RSS, 12/23RSS 12/12RSS and 23/23RSS). All DNA substrates used in TPM have universal biotin 5-AAAGGGAATAAGGGCGACAC-3 and digoxigenin 5-TATGGAAAAACGCCAGCAACG-3 labeled primers.

3.5.4 *In vitro* cleavage assay

The oligonucleotide RAG cleavage assay detects nicked and hairpin products and was performed as previously described [14]. First 125nM RAG1/2c or RAG1(D708A)/2c and 560nM HMGB1 were preassembled in an 8μL reaction in dilute dialysis buffer (DDB: 25mM Tris-HCl (pH: 8.0), 150mM KCl, 2mM DTT) and 0.1% glycerol for 5 min. Then 12μL of EMSA reaction buffer (ERB: 25mM Tris-HCl (pH: 7.4), 65mM potassium acetate, 5mM MgCl2, 100μg/mL acetylated BSA (Sigma-Aldrich) and 10% DMSO) containing 2nM 12RSS (hot) and 8nM 23RSS (cold) oligonucleotides substrates was added to RAG1/2c and HMGB1 bringing the DDB and ERB together to form the RAG reaction buffer (RRB), and the reaction volume to 20μL with 50nM RAG1/2c, 225nM HMGB1, 1nM 12RSS and 4nM 23RSS. The reactions were incubated at 37 °C for 1hr, stopped with 10mM EDTA in 95% formamide and incubated at 95 °C for 2 min. The reaction products were fractionated on a 15% 19:1 polyacrylamide sequencing gel (National Diagnostics) and visualized with a phosphoimager.

3.5.5 Tethered particle motion: Flow cell assembly, data acquisition, data analysis and calibration

TPM flow cells assembly begins by first drilling 12 holes into a pre-cleaned glass slide with a 1.5mm diamond tipped drill bit (CRLaurence) to create inputs and outputs for six channels. The drilled glass slide (Corning) and a partner coverslip (Fisher) are son-

icated (Aquasonic-Model 75T) in 100% EtOH for 30 min, rinsed with ddH_2O, sonicated in 1M KOH for 30 min, rinsed with ddH_2O, sonicated in 5% HCl, rinsed with ddH2O, and functionalized with epoxysilane (116mL Isopropanol (IPA), 1mL ddH_2O, 3mL of 3-glycidoxypropyldimethoxymethylsilane (Sigma) and 120 μL N,N-Dimethylbenzylamine (Sigma)) at RT for 1 h and 30 min. The drilled glass slides and coverslips were washed in IPA, dried with N_2 gas and baked at 100C for 30 min. Next 0.02 in. ID/0.06 in. OD tygon microbore tubing (Cole-Parmer) is placed in each of the 12 holes of the drilled glass slide, designating an input and output for all six channels. Next five-minute epoxy (Devcon) is applied at the tubing-glass interface, and a piece of double-sided tape (0.12mm thick, Grace Bio-Labs) with six holes designating the channels of the flow cell is placed on the drilled glass slide and is sandwiched between a coverslip. After the flow cell is assembled polyclonal sheep anti-digoxigenin (Roche) is added to each of the six channels to a final concentration of 40ng/μL. The anti-digoxigenin (anti-dig) solutiion is incubated overnight at RT to allow covalent coupling of anti-dig to epoxy groups within the channels of the flow cell. Next, TPM assembly buffer (TPM-AB: 20mM Tris-Acetic Acid pH 8.0, 130mM KCl, 1mM MgCl2, 0.1mM DTT, 0.1mM EDTA, 20μg/mL Aceylated-BSA and 3mg/ml Biotin-free Casein) is added to the flow cells channels and incubated for 1h in order to flush out uncoupled anti-digoxigenin and quench active epoxy groups. Then biotin and digoxigenin labeled DNA is added to each channel to a final concentration of 5pg/μL and incubated for 20 min. The excess DNA is washed away with TPM-AB and 0.49μm streptavidin coated polystyrene beads (Bangs Laboratories) diluted in TPM-AB to a final concentration 3 X 10^{10} beads/mL are added to each channel and incubated for 10 min. The excess beads are washed away with TPM-AB and RRB completing the assembly of DNA-tethered beads. The DNA tethered beads were imaged using brightfield microscopy (because the polystyrene beads scatter light) on inverted Olympus IX71 microscopes with either a 100x oil objective, or a 60x oil objective with a 1.6x magnifier. A Basler A602f camera was used to acquire images at 30Hz. We track DNA-tethered beads by cross-correlating each frame with the initial frame in a time series for each bead. This generates raw x and y positions of each beads excursion, relative to its anchor point. Drift was corrected from these raw data by subtracting the results of a low-pass first-order Butterworth filter with a cutoff frequency $f_{cB} = 0.05$ Hz. The root-mean-square (RMS) motion was obtained by applying a Gaussian filter with a -3 dB frequency of $f_{cG} = 0.0326$ Hz corresponding to a 8 second or 240 frame standard deviation

of the filter, to the mean-squared displacement of the data ($\overrightarrow{x}^2 + \overrightarrow{y}^2$). The RMS motion of the bead is the square root of the result of the convolution of ($\overrightarrow{x}^2 + \overrightarrow{y}^2$) and the Gaussian filter. We calibrate TPM by tracking the RMS of DNA molecules of various lengths (252, 539, 736, 946, 1124, 1316, 1521, 1717, 1910, 2077, 2280 and 2900bp) and perform a quadratic fit to data to extract an equation that relates RMS to DNA length in bps. To define a good DNA tethered bead we used symmetry-of-motion and length-of-motion checks used previously [15]. We then track the good beads in RRB, then we stop tracking the beads, and add a given amount of RAG1/2c or RAG1/2c-HMGB1 and continue tracking the same beads. This scheme allows for precision a measurements of changes to the RMS motion of DNA-tethered beads by RAG1/2c and RAG1/2c-HMGB1. Bad beads who passed the initial checks are discarded if the beads stick to the surface and if the beads initial RMS is not within error to the previously measured RMS for a particular DNA length.

3.5.6 Statistical Mechanics of the RAG1/2c

We fit the single molecule RAG trajectories to two Gaussians to term the probability of RAG1/2c being in the bound or unbound state. Then we built a statistical mechanical model using Boltzmann's distribution, to compute p_{bound} and $p_{unbound}$ and fit the equation for p_{bound}, to the data.

$$p_{unbound[R(0)]} = \frac{1}{1 + \frac{R}{R_o} e^{-\beta \triangle E_{12or23RSS}}}$$

$$p_{bound[R(1)]} = \frac{\frac{R}{R_0} e^{-\beta \triangle E_{12or23RSS}}}{1 + \frac{R}{R_o} e^{-\beta \triangle E_{12or23RSS}}}$$

Where $\beta = 1/\mathrm{K}_B\mathrm{T}$ R = RAG1/2c concentration, R_o = 1M, $\triangle\ \mathrm{E}_{12RSS}$ = -18.09K_BT and $\triangle\ \mathrm{E}_{23RSS}$ = -16.93K_BT. In the presence of HMGB1 $\triangle\ \mathrm{E}_{12RSS}$ = -19.31K_BT and $\triangle\ \mathrm{E}_{23RSS}$ = -18.54K_BT

3.5.7 Kinetic Theory of the RAG-RSS binding

To analyze the transition rates and dwell times in TPM data, bead trajectories were first thresholded based on RMS value to be in either a bound (smaller RMS) or unbound (higher RMS) state or one of two spurious states; the first spurious state corresponds to

transient sticking event where the RMS of the bead plummets as its movement is halted and the second state occurs when the tracking program has a tracking error and the RMS of the bead jumps dramatically. To identify these two spurious states we also applied an upper and lower threshold, these thresholds were chosen conservatively and were unambiguous, both sticking and tracking errors produced enormous changes in the RMS of the bead compared to the small changes observed from protein binding/unbinding. The threshold between the bound and unbound state was chosen by eye from a histogram of RMS values in a sample where the bead spends roughly half of its time in either of the two states (bound and unbound).

Next, to reduce the influence of noise in the trajectories, short states were removed. This corresponded to removing states that did not last longer than four times the dead time of the Gaussian filter applied to the data, corresponding to roughly 21 seconds in this analysis (3). These states were removed by assigning the first half of the spurious state to the previous state of the bead and the second half of the spurious state to the next state of the bead. If a short state began (or ended) a trace, the entire duration was assigned to the next (previous) state. Spurious states were handled in a similar manner; spurious states occurring in the middle of the trace were assigned in the same way (the first half to the previous state, the second half to the next state) and spurious states that began a trace were assigned to the next state. However, spurious states that ended a trace were simply discarded. Statistics were then gathered on the traces produced. When calculating the mean dwell time in the bound state (Figure 3.5) we applied a correction, which accounts for the fact that even in the absence of protein, the bead occasionally dipped below the threshold and gave a spurious on state. To correct for this, we calculate the probability per unit time of such a transition, ρ_{false}, from our collection of no protein traces by counting the total number of false transitions observed divided by the total time of calibration across all observed beads. We also calculate the mean dwell time of one of these false transitions, τ_{false}. From here we can calculate the expected number of false transitions observed, n_{false}, for each concentration of RAG1/2, which is ρ_{false} times the total time the bead is observed in the off state. The measured mean dwell time in the on state $< \tau_{\text{obs}} >= \omega < \tau_{\text{on}} > +(1-\omega) < \tau_{\text{false}} >$ where ω is the fraction of transition events measured that are expected to be real events (total number subtracted by n_{false} divided by

the total number). Thus, the mean dwell time in the on state is calculated by,

$$< \tau_{\text{on}} >= \frac{< \tau_{\text{obs}} > -(1-\omega) < \tau_{\text{false}} >}{\omega}$$

Similarly, this correction was done to the off-time measurement, $<\tau$ unbound>. We calculated the average time in the off state by taking the total time we measured the bead in that state divided by the total number of transitions observed in that time. We correct for false transitions in this data by removing the expected number of false transitions from the total number and adding the expected total time of those transitions to the total time. The dwell time histograms of Figure S3 have not had the above correction applied.

3.5.8 TPM: RAG1/2c-RSS, RAG1/2c-HMGB1-RSS binding, hairpin production and paired complex detection assays

Single-molecule experiments performed at saturating concentrations of RAG1/2c were assembled by first diluting the RAG proteins to 200nM in 50μL of DDB then 50μL of ERB was added followed by 100μL of RRB yielding a 200μL final volume of 50nM RAG1/2c which was added to the TPM flow cell by gravity flow. The RMS motion of the DNA tethered beads was tracked for 700 sec before and after RAG1/2c was added. To acquire the experimental data for determining dissociation constants, RAG1/2c was titrated (1, 2.5, 5, 10, 25 and 50nM) in the presence of a single RSS and the DNA tethered beads were tracked for 4,000 sec. RAG1/2c-HMGB1 experiments performed at saturation first preincubated 50nM RAG1/2c and 25nM HMGB1 in DDB for 10 min then the proteins were diluted in RRB to a 200μL final volume and added to DNA tethered beads. The RMS motion of the DNA tethered beads was tracked for 700 sec before and after the addition of RAG1/2c-HMGB1. The hairpin production assay was performed from t = 0h to t = 2h in which we counted the beads initially at t = 0h to ensure we started with 70-80 beads. Next we added RAG1/2c and HMGB1 and monitored bead loss after t = 2h or every 10 min for t = 2h to capture the dynamics of hairpin production. The % hairpin production for the 12/23RSS was determined using the equation $\triangle$NBR/Ntot = $\triangle$Ntot/Ntot - $\triangle$NPBR/Ntot where N is number of beads, NBR is number of beads released, NPBR is the number of beads released passively, and Ntot is the total number of beads. The % non-specific DNA compaction is derived from the change in <RMS> at different HMGB1 concentrations

on the NoRSS DNA substrate..Paired complex detection assays where performed by track DNA tethers with a 2900bp contour and 1200bp or 1800bp intersignal spacing for 1h, in the presence of 5nM or 25nM RAG and 80nM HMGB1. The trajectories resulting in bead loss were inspected for a signature of paired complex formation. Single molecule binding, hairpin production, and paired complex detection assays were conducted at RT.

Single molecule experiments performed at saturating concentrations of RAG1/2c were assembled by first diluting the RAG proteins to 200nM in 50μL of DDB then 50μL of ERB was added followed by 100μl of RRB yielding a 200μL final volume of 50nM RAG1/2c which was added to the TPM flow cell by gravity flow. The RMS motion of the DNA tethered beads was tracked for 700 sec before and after RAG1/2c was added. To acquire the experimental data for determining dissociation constants RAG1/2c was titrated (1, 2.5, 5, 10, 25 and 50nM) in the presence of a single RSS and the DNA tethered beads were tracked for 4,000 sec. RAG1/2c-HMGB1 experiments performed at saturation first preincubated 50nM RAG1/2c and 25nM HMGB1 in DDB for 10 min then the proteins were diluted in RRB to a 200μL final volume and added to DNA tethered beads. The RMS motion of the DNA tethered beads was tracked for 700 sec before and after the addition of RAG1/2c-HMGB1.

3.6 References

[1]. Swanson P and Schatz D, (2011) V(D)J recombination: mechanisms of initiation. *Annu Rev Genetics*45:167-202

[2]. Lee, G.S. *et al.*, (2004). RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* 117, 171-184

[3]. Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 71, 101-132.

[4]. Jones, J.M., and Simkus, C. (2009). The roles of the RAG1 and RAG2 "non-core" regions in V(D)J recombination and lymphocyte development. *Arch Immunol Ther Exp (Warsz)* 57, 105-116.

[5]. Swanson, P.C. (2004). The bounty of RAGs: recombination signal complexes and reaction outcomes. *Immunol Rev* 200, 90-114.

[6]. van Gent DC *et al.* (1997) Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO* 16(10):2665-70.

[7]. Swanson PC, (2002) A RAG1/RAG2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals.*Mol. Cell. Biol.* 22(22):7790-801

[8]. Bergeron S *et al.*, (2005) Both high mobility group (HMG)-boxes and the acidic tail of HMGB1 regulate recombination-activating gene (RAG)-mediated recombination signal synapsis and cleavage in vitro. *J. Biol Chem.* 280(35):31314:24

[9]. Kriatchko AN *et al.*, (2008) HMG-box domain stimulation of RAG1/2 cleavage activity is metabl ion dependent. *BMC Mol Biol.* 9:32

[10]. Zhang M and Swanson PC, (2009) HMGB1/2 can target DNA for illegitimate cleavage by the RAG1/2 complex. *BMC Mol Biol.* 10:24

[11]. Lewis, S.M. (1994). The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Adv Immunol* 56, 27-150.

[12]. Kumar S and Swanson PC (2009) Full-length RAG1 promotes contact with coding and intersignal sequences in RAG protein complexes bound to recombination signals paired in cis. *NAR* 37(7):2211-26.

[13]. Altschul SF *et al.*, (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410

[14]. Finzi L and Gelles J (1995) Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules. *Science* 267(5196):378-80

[15]. Manzo C *et al.*, (2012) The effect to nonspecific binding of lambda repressor on DNA looping dynamics. *Biophys J.* 103(8):1753-61.

[16]. Mumm JP *et al.*, (2006) Viewing single lambda site-specific recombination events from start to finish. *EMBO J.* 25(19):4586-95.

[17]. Pouget N *et al.*, (2006) IS911 transpososome assembly as analysed by tethered particle motion. *NAR* 34(16):4313-23.

[18]. Pavlicek JW *et al.* (2008) Quantitative analyses of RAG-RSS interactions and conformations revealed by atomic force microscopy. *Biochemistry* 47(43):11204-11

[19]. Han L *et al.*, (2009) Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One* 4(5):e5621.

[20]. Johnson SJ *et al.*, (2012) Sequence dependence of transcription factor-mediated DNA looping. *NAR* 40(16):7728-38.

[21]. Wang G *et al.* (2012) Real-time monitoring of RAG-catalyzed DNA cleavage unveils

dynamic changes in coding end association with the coding end complex. *NAR* 40(13):6082-96.

?[22]. Zhao, S. *et al.*, (2009). A non-sequence-specific DNA binding mode of RAG1 is inhibited by RAG2. *J Mol Biol* 387, 744-758.

[23]. McCauley MJ. *et al.*, (2007). HMGB binding to DNA: single and double box motifs. *J Mol Biol* 374(4):993-1004.

[24]. Shlyakhtenko LS *et al.*, (2009). Molecular mechanism underlying RAG1/RAG2 synaptic complex formation. *J Biol Chem.* 284(31);20956-65

[25]. Ciubotaru M *et al.* (2013) RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. *NAR* 41(4):2437-54.

[26]. Askary A *et al.*, (2014) Modeling of the RAG reaction mechanism. *Cell Reports* 7(2):307-15

[27]. Sheehan, K. M., and M. R. Lieber. (1993) V(D)J recombination: signal and coding joint resolution are uncoupled and depend on parallel synapsis of the sites. *Mol. Cell. Biol.* 13:1363-1370.

[28]. Eastman, Q. M. *et al.* (1996) Initiation of V(D)J recombination in vitro obeying the 12/23 rule. *Nature* 380:85-88.

[29]. van Gent DC *et al.*, (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* 85(1):107-13

[30]. Ji Y *et al.*, (2010) The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* 141(3): 419-31

3.7 Appendix 1: Supplementary Information

3.7.1 Calibrating the tethered particle motion (TPM) assay.

To calibrate the TPM assay we generated various DNA substrates ranging in length from 252-2900bp and tethered them to beads (Figure 3.2A) and tracked their <RMS> motion (Figure 3.2B). We then performed a quadratic fit to the data for <RMS> as a function of DNA length (Figure 3.2C) and extracted an equation that allows for use to interconvert between <RMS> and DNA length.

3.7.2 Representative RAG1/2c trajectories, normalized probability distribution, and absolute reduction in DNA length by RAG1c, RAG2c, RAG1/2c and RAG mutants

To determine the mean dwell times for RAG1/2c shown in Figure we computed the time spent in the unbound and bound states over a range of concentrations with representative trajectories shown in Figure 3.4A,B. We added the highest concentration of RAG1/2c used in the titration to DNA substrates with no, one or two RSS (Figure 3.4C) and found that nearly all substrates were bound, which is shown in the composite trajectories and normalized probability distributions (see Figure 3.4D-I). We also determined if RAG1/2c alone formed a paired complex using an intersignal spacing of 250bp between two RSSs, and only observed a 100bp reduction in DNA length (Figure 3.4G-J). This is two times the reduction in the tether length for a single RSS but two-fold less than the intersignal distance indicating we are observing RAG1/2c complexes bound to two RSSs on a single DNA substrate. In Figure 3.4J we show data illustrating the detection of individual RAGs and RAG mutants binding to one or two RSSs. We find that at 50nM, RAG1c can bind RSSs but also exhibits promiscuous binding as evidenced by the size of the error bars relative to RAG1/2c (Figure 3.4J). Furthermore, we show RAG1/2c in the presence of different divalent cations, RAG1/2c catalytic mutants and MBP free RAG1/2c can shorten DNA in presence of RSSs (Figure 3.4J). However, RAG2c and RAG1/2c lacking a nonamer binding domain (RAG1(No-NBD)/2c) do not bind either RSS.

3.7.3 Using Waiting times to compute the dissociation constant

As noted in the main text, there are several different ways of obtaining dissociation constants which in principle should give the same results. The main text used Gaussian fits to the main peaks. Alternatively, one can compute the fraction of time that the system spends in the bound state as a function of concentration and this should yield a fit to the dissociation constant. In Figure 3.6, we show the data used to carry out this analysis. Figure 3.6A shows the dwell times in both the bound and unbound states as a function of RAG1/2c concentration. In Figure 3.6B we plot the fraction of time RAG1/2c is bound (solid points) as a function of concentration in the presence of 12RSS (blue points) or 23RSS (red points). This is calculated by taking the fraction of time the system spends in the bound state versus

the total time of the trajectory of interest. Each binding probability curve (12RSS or 23RSS) is fit for Kd to a Langmuir isotherm (equivalent to a Hill function with n=1), which follows directly from the chemical equation for equilibrium binding/unbinding of a molecule to a target. We determined a Kd = 18.5nM ± 2.9nM for the 12RSS which is within error of the Kd determined using the other analysis method, Kd = 13.9 ± 4.7nM for RAG1/2c-12RSS, and which is in reasonable accord with both previously reported literature values of 14.5nM ± 2.4nM (21) and 25.0nM ± 5.0nM (22) (Figure 3.6). In addition we found Kd = 54.7nM ± 1.1nM for the 23RSS which is larger than the Kd = 44.4nM ± 6.5nM obtained using Gaussian fits to the bulk binding probability.

3.7.4 Expected dwell time in off state from simple diffusion

The flux of particles arriving at a target per unit time is

$$\frac{dn}{dt} = 4\pi D\alpha c$$

where α is the size of the absorbing target (the RSS site in this case), c is the concentration of the absorbing molecule (RAG1/2c in this case) and D is the diffusion constant of the absorbing molecule. The diffusion constant of a particle in a solvent can be written in terms of temperature (T), the solvent viscosity (ν) and the particle size,

$$D = \frac{k_B T}{6\pi\nu d}$$

where kB is Boltzmanns constant. At room temperature, k_B =4.11x10-21J and $\nu 10^3$ Pa s for water at room temperature. The hydrodynamic radius of RAG1/2c is roughly 10nm and we will assume the size of the target is roughly 1nm. So we conclude that,

$$\frac{dn}{dt} \approx 0.1 \left[\frac{particles}{s * nM}\right]$$

However, for our data the leading constant is (2.5±0.5)x10^{-4}. A difference of 3 orders of magnitude. However, we note that this calculation is made in the limit of a perfect absorber meaning that every time a molecule arrives at the RSS it will bind. This completely neglects factors such as the orientational effect which requires the protein to arrive at its binding site in the correct orientation.

3.7.5 Effect of DNA length flanking RSS sites on RAG-HMGB1-RSS complex formation and percent DNA compaction by HMGB1 as a function of concentration for a range of DNA lengths.

As noted in the main text, one concern in performing these TPM experiments is the proximity of the bead and the coverslip to the relevant RSSs. To explore the consequences of the bead for the protein-DNA complexes, we performed a series of experiments in which the distance between the RSS and the bead was varied as shown in Figure 3.7A. Indeed, we find that the distance between the RSS and the bead has to be sufficiently large in order to avoid restricted motion. As shown in Figure 3.7B, increasing the distance between the bead and the RSS from 126bp to 426bp leads to a substantial increase in the RMS motion. We also find the increase in RMS motion is conserved for RAG1/2c and catalytically inactive RAG1(D708A)/2c in the presence of HMGB1, indicating the reduction in length is independent of RAG nicking activity, and by inference is caused by HMGB1 (Figure 3.7B). We also determined the percent DNA compaction by HMGB1 alone from 5nM-225nM for DNA substrates ranging in length from 539-2280bp. We found the percent DNA compaction was relatively the same at a particular concentration of HMGB1(Figure 3.7C).

3.7.6 Composite trajectories and normalized probability distributions of NoRSS and 12RSS bead loss, and 1200bp and 1800bp bead loss events preceded by paired formation.

We performed experiments to directly observe the paired complex on DNA substrates with an intersignal spacing of 1200bp and 1800bp shown in Figure 3.8. To confirm that the paired complexes resulting in bead loss were dependent on two RSSs we performed experiments in the presence of NoRSS or 12RSS and did not observe paired complex formation prior to bead release (Figure 3.9A,B). We determined the mean dwell time of the trajectories resulting in paired complex formation prior to bead release on DNA substrates with an intersignal spacing of 1200bp or 1800bp represented at black bars shown in Figure 3.8C. All the trajectories that went into those two dwell time distributions are shown in Figure 3.9C,D.

3.7.7 Catalytically active RAG proteins are required for bead release

To determine that bead release was dependent on catalytically active RAG we first performed a bulk hairpin production assay and found we only observed hairpin production when using RAG1/2c and not catalytically inactive RAG1(D708A)/2c (Figure 3.11A). After establishing a single-molecule hairpin production assay that uses bead release as a readout, we performed another useful control to further confirm and clarify that bead loss is indicative of the actual RAG-HMGB1 dynamics. In particular, we used mutant proteins to confirm catalytically active RAG1/2c was responsible for bead release. To that end, we added 50nM RAG1/2c or catalytically inactive RAG1(D708A)/2c and 500nM HMGB1 to DNA tethers with a 12 and 23RSS. We only observe bead release when using RAG1/2c but not when using RAG1(D708A)/2c (Figure 3.11B,C). As an indirect control to determine that synapsis preceded hairpin production represented as bead release in our assay we reduced the intersignal spacing from 1000bp to 73bp (Figure 3.11D). The 73bp intersignal spacing was previously shown to ablate hairpin production in a bulk assay, and we find that we also hamper bead release in our single molecule assay (Figure 3.11E). Interestingly we can rescue bead release by substituting Mn2+ for Mg2+, which eliminates the requirement for synapsis to precede hairpin production (Figure 3.11F).

3.8 Appendix 2: TPM Trajectories and Distributions

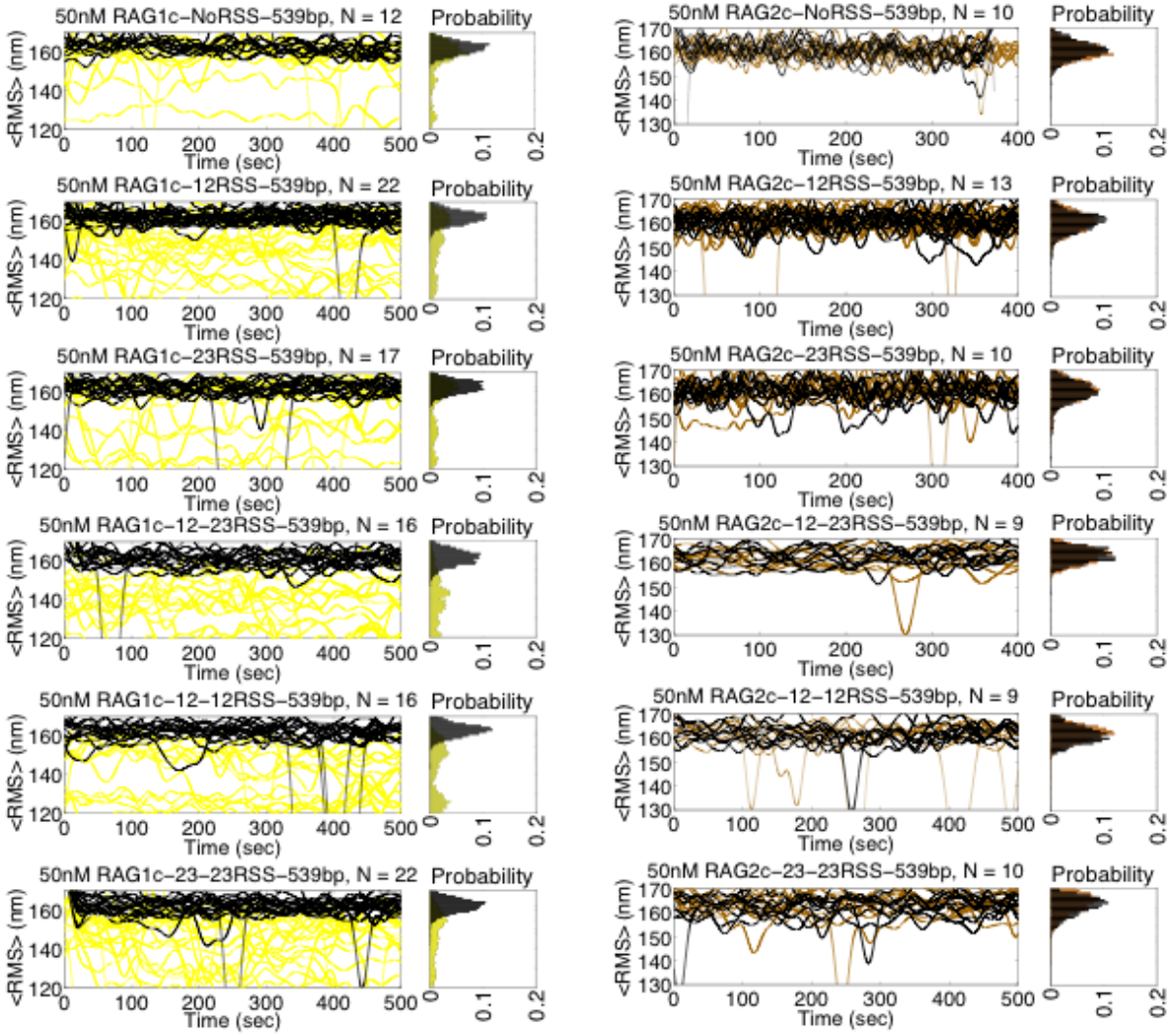

Figure 3.13: RAG1c and RAG2c

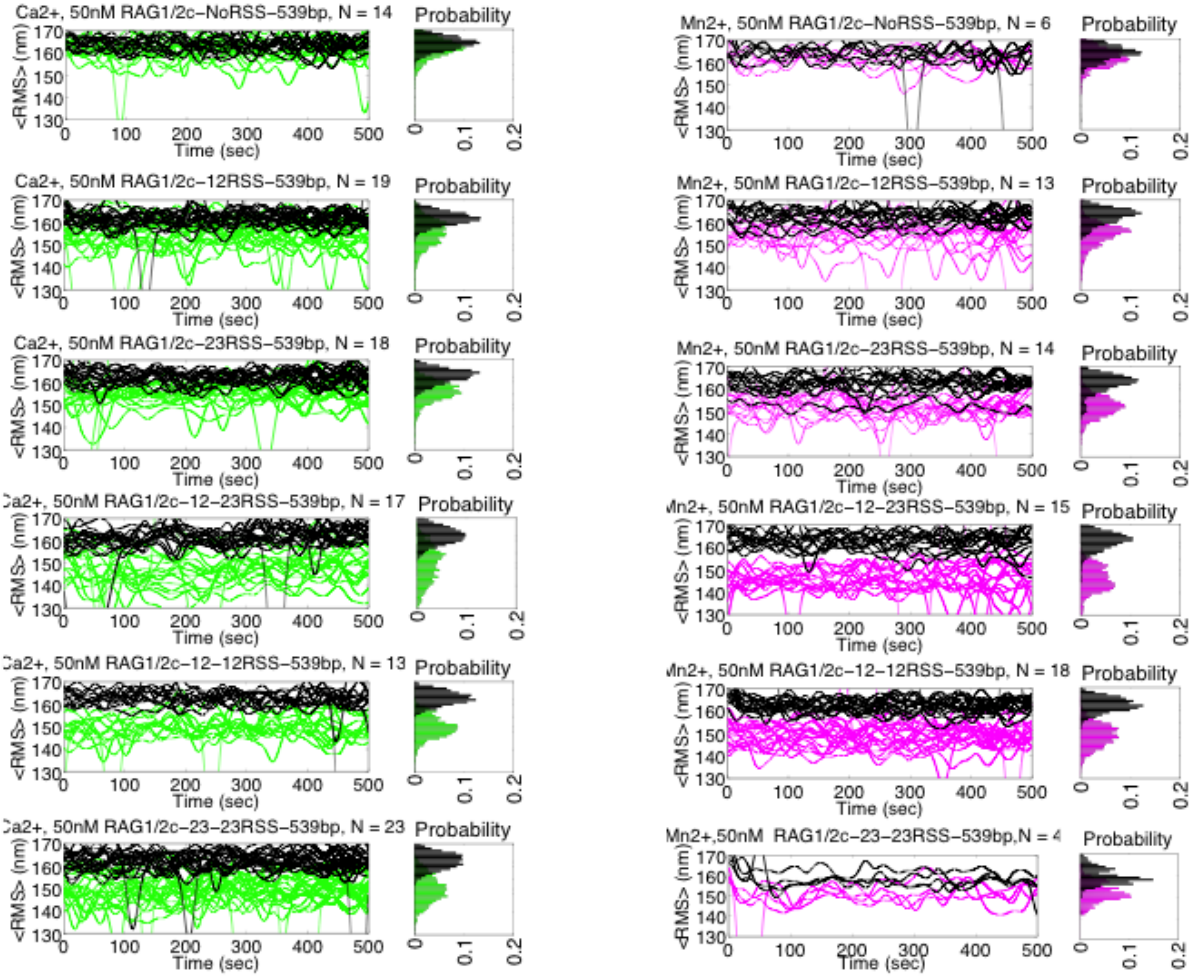

Figure 3.14: RAG1/2c-Ca2+ and RAG1/2c-Mn2+

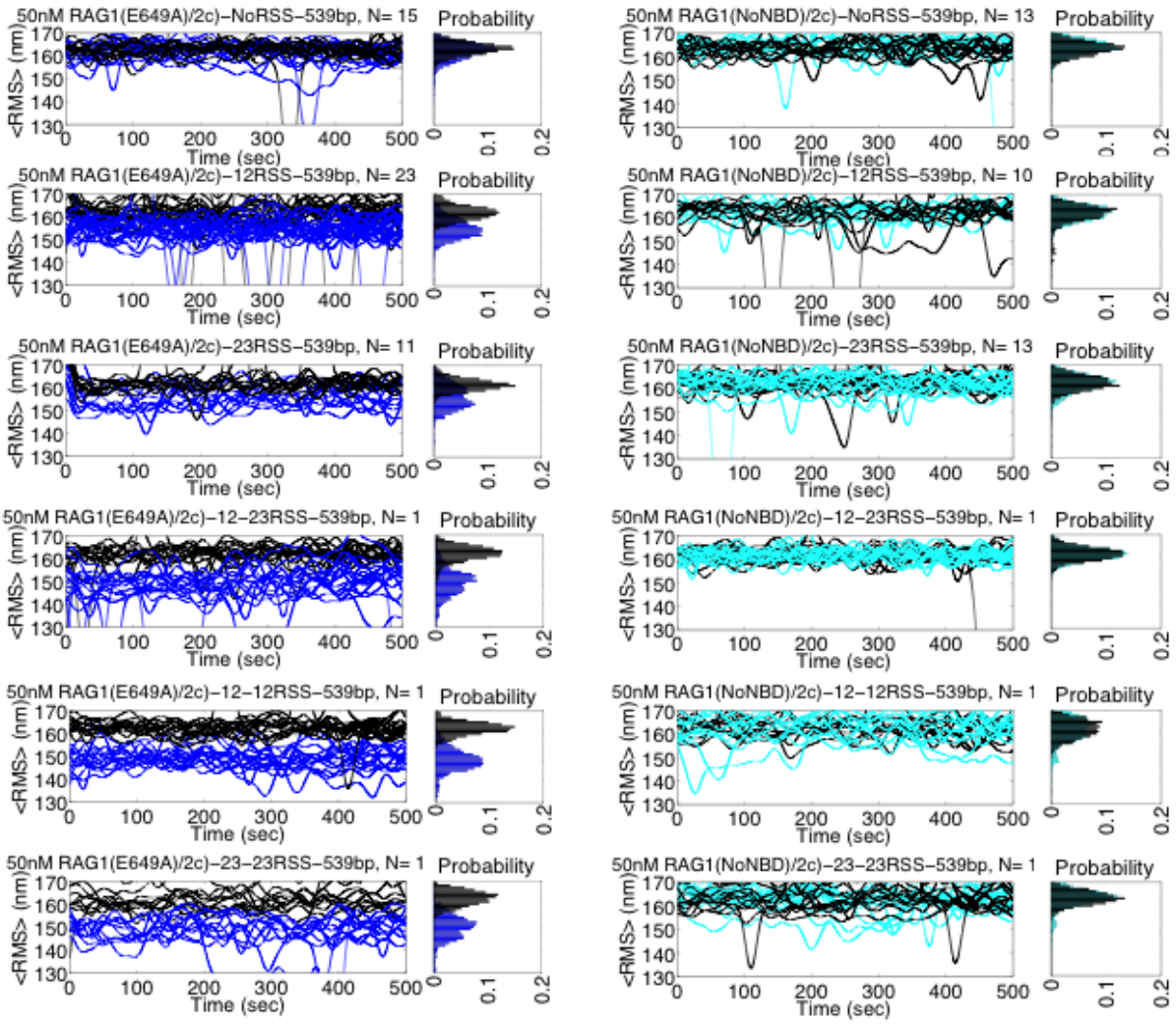

Figure 3.15: RAG Mutants

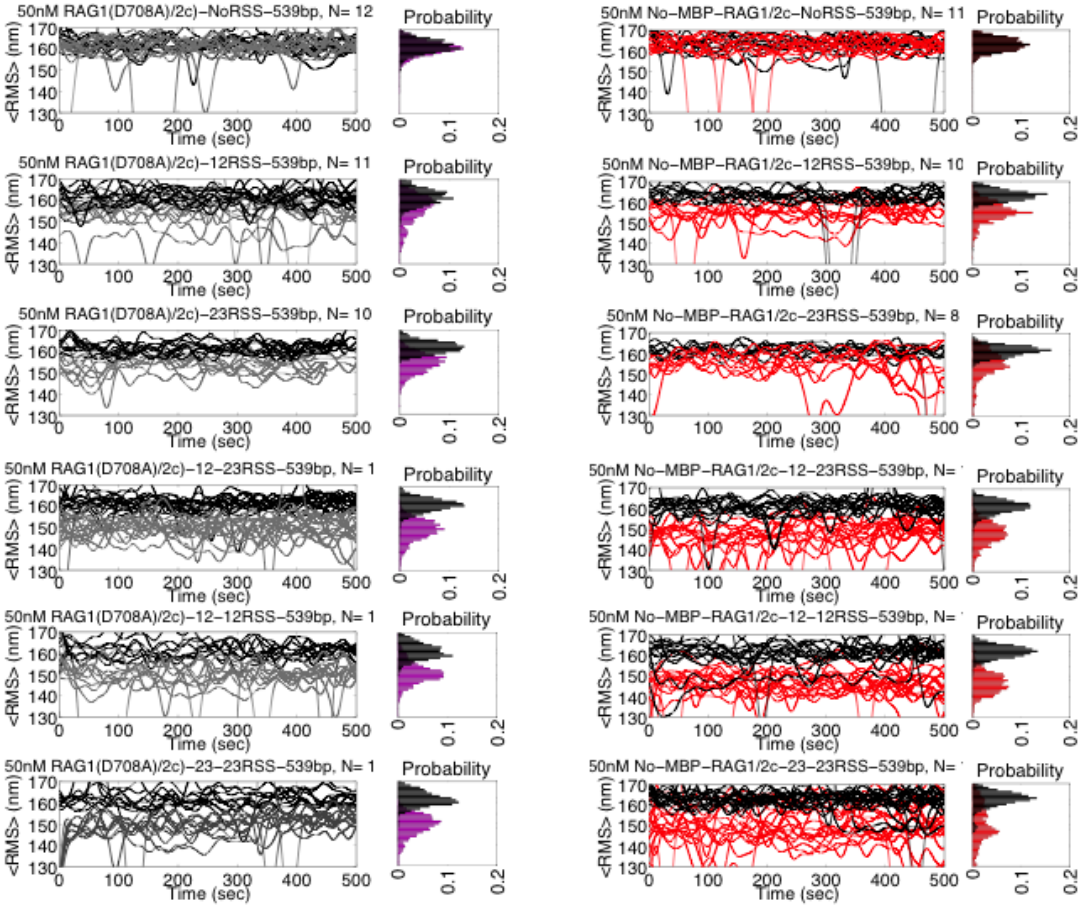

Figure 3.16: RAG Mutants

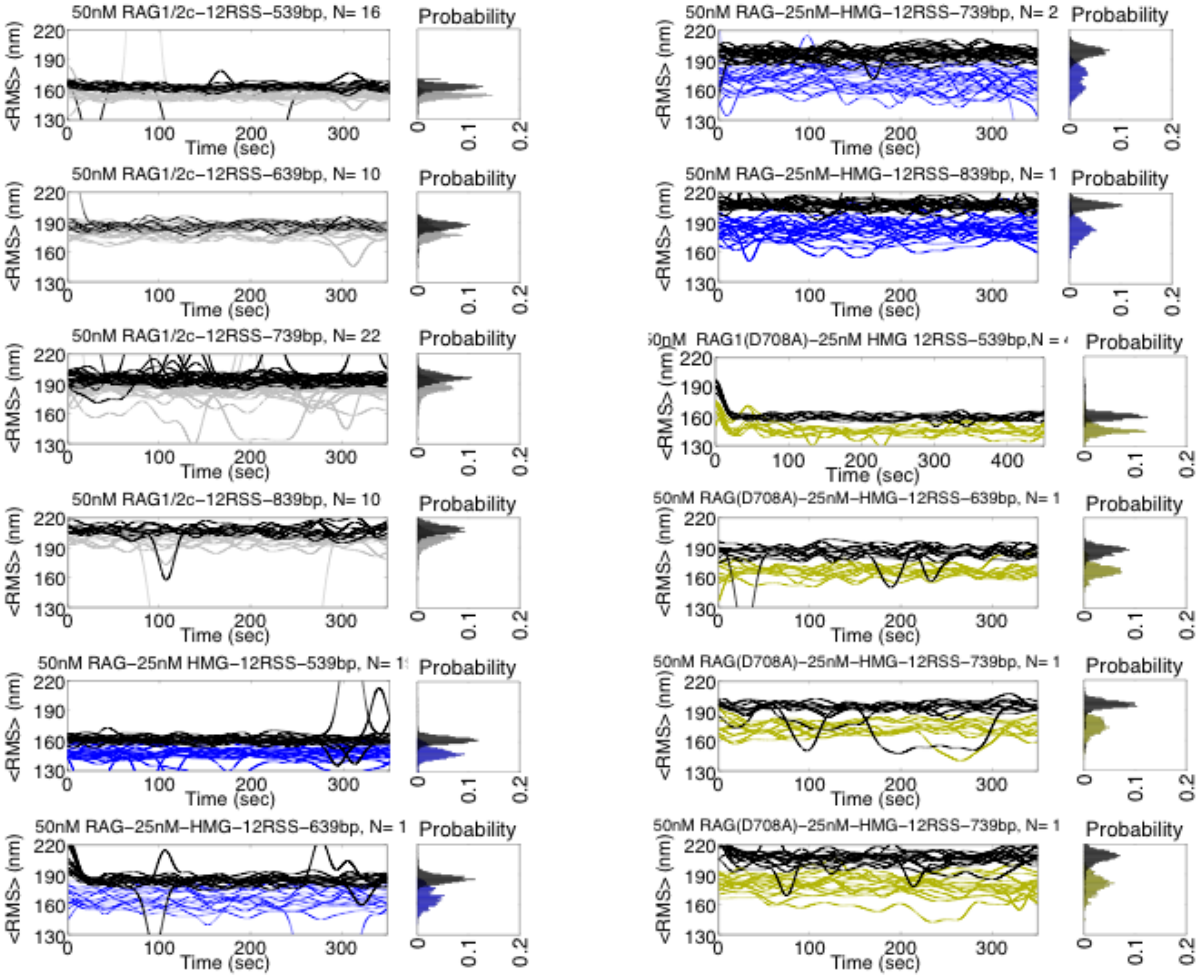

Figure 3.17: RAG and HMGB1 length dependence

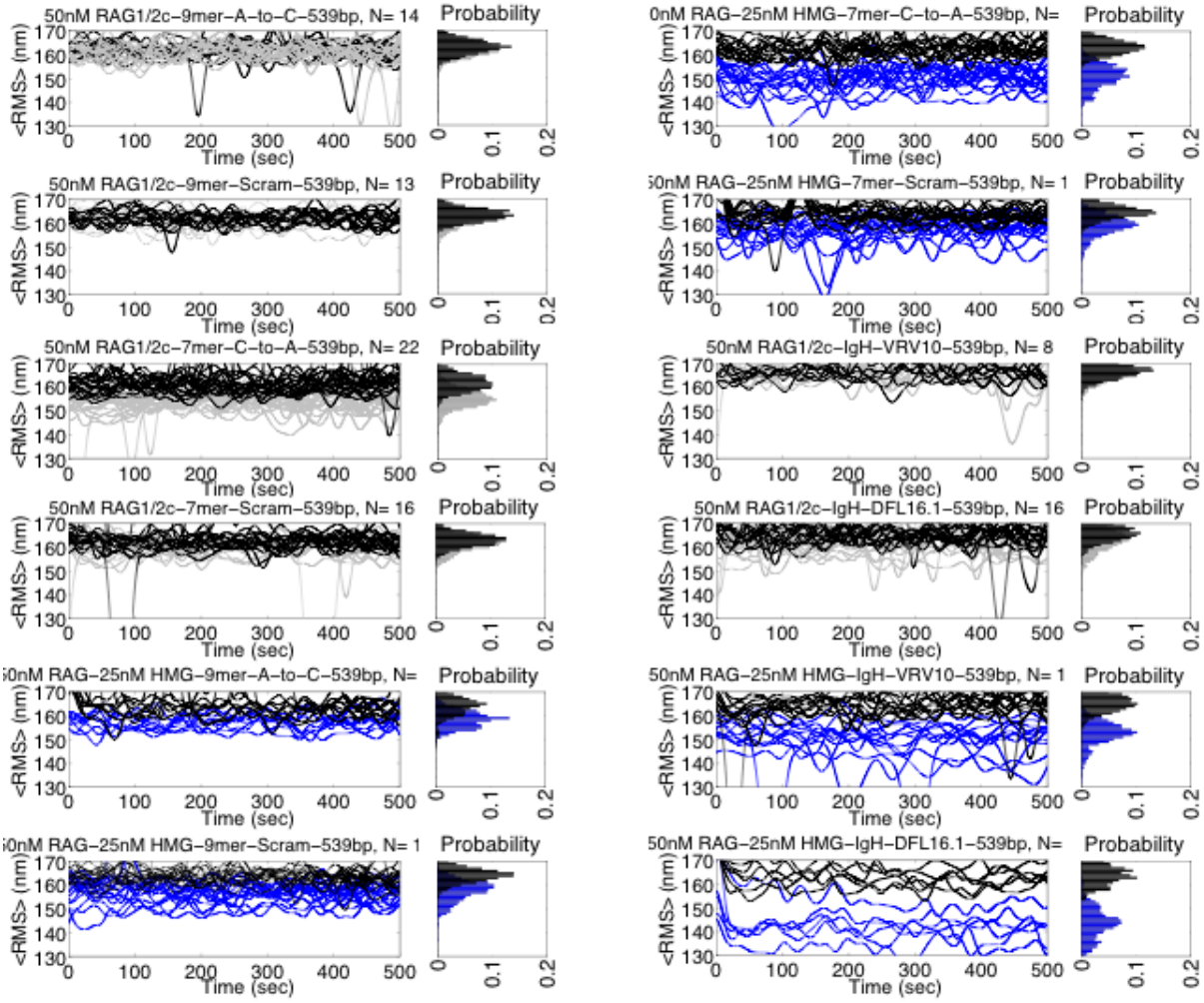

Figure 3.18: RSS mutants and non-consensus RSSs

Part II

Genome Packaging

Chapter 4

Fractional HMGB1-DNA compaction is conserved across DNA lengths and modulated by divalent cations.

4.1 Abstract

Eukaryotic high-mobility group box (HMGB) proteins bind non-specifically to DNA and alter its flexibility, as well as modulate chromatin, transcription, replication, repair, and recombination. HMGB1 bends DNA locally, which results in DNA compaction. Single molecule studies have revealed the binding properties of HMGB1, but all techniques used have applied force. Here, we conduct single molecule measurements of HMGB1 binding behavior under thermal forces. We observed concentration dependent DNA compaction, differential compaction depending on the divalent cation type, and determine the fraction of DNA compacted at a particular HMGB1 concentration is independent of DNA length. This study furthers our understanding of HMGB1-mediated DNA compaction.

4.2 Introduction

The high mobility group box (HMGB) proteins are critical in chromatin regulation [1-4]. Their name is derived from the fast migration observed once the proteins undergo native gel electrophoresis [2]. HMGB proteins are a family of non-specific DNA binding proteins that are conserved in eukaryotes, and are the second most ubiquitous DNA binding proteins in the nucleus next to histones. HMGB proteins can also act as chaperones to facilitate DNA binding of tumor suppressors, transcription factors, histones, and the recombination activating genes through indirect or direct interactions, and as a result can modulate genome management processes in eukaryotes [5-8].

An HMG box domain is made up of 80 amino-acids [1]. The dominant HMGB variant HMGB1 has two HMG box domains coupled by a flexible linker, and has been shown to bind and bend the DNA minor groove non-specifically [3]. Single molecule studies using optical tweezers and atomic force microscopy have shown that HMGB1 reduces the persistence length of DNA by introducing a $67\,^{\circ}\pm1.3\,^{\circ}$ bend [9,10]. The bend has been shown to behave as an intermediate between a static kink and a flexible hinge [10]. HMGB1 is conserved in most eukaryotes, and so are its molecular properties, but genome size can vary over several orders of magnitude across eukaryotes. Here we investigate how DNA length effects HMGB1-mediated DNA compaction. We show HMGB1 compacts DNA in a concentration dependent manner, find that the divalent cation type can alter the amount of DNA compacted, and reveal that HMGB1 compacts the same fraction of DNA independent of contour length.

4.3 Results

4.3.1 HMGB1 reduces the contour length of DNA

First, we measured HMGB1-mediated DNA compaction using tethered particle motion (TPM), a technique from single molecule biophysics that provides a direct window into the behavior of DNA binding proteins under thermal forces (Fig 4.1a) [11]. We assembled 539bp DNA tethered beads and titrated 5, 10, 25, 50, 112, 225, and 1000nM of HMGB1. We found that HMGB1 reduces the <RMS> or end-to-end distance of the DNA tethered beads in a concentration dependent manner (Fig 4.1b, c). To determine the fraction of

DNA compacted, we first converted the change in <RMS> to a change in kilo base pairs (kbp), then divide the absolute change in DNA length in kbp by the DNA contour length in kbp(Fig 4.1d, e). The concentration dependent reduction in DNA length is consistent with what others have observed using other single molecule techniques [9,10]. Now that we can measure HMGB1-mediated DNA compaction, we investigated the effect of changing the divalent cations in the binding buffer.

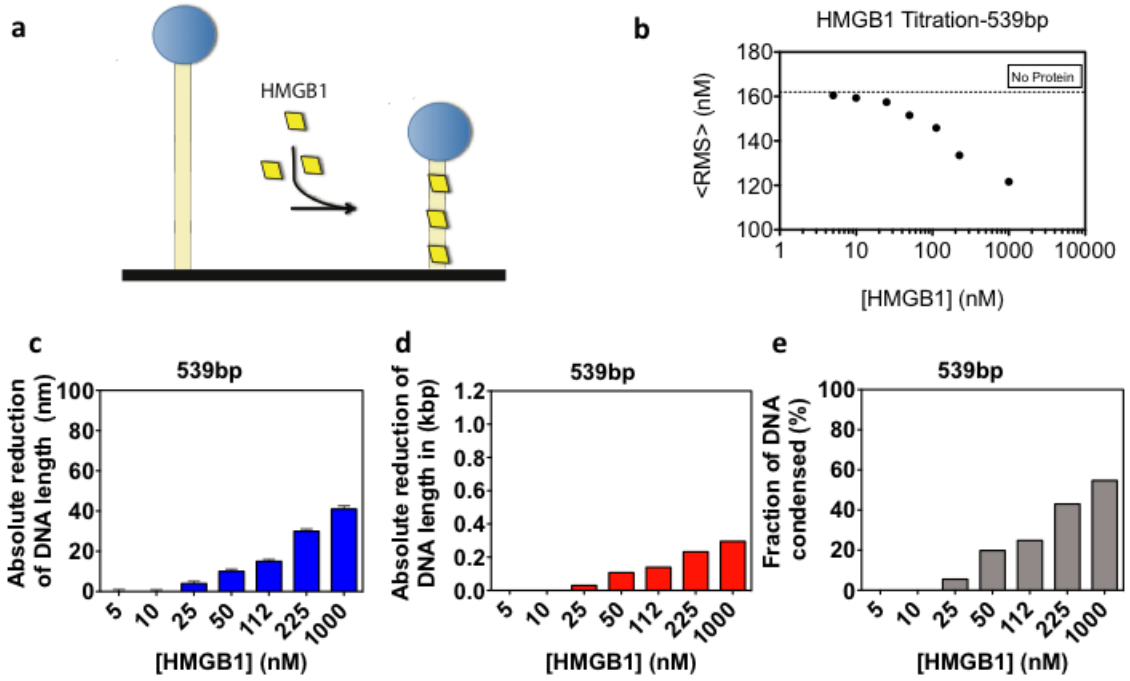

Figure 4.1: HMGB1 reduces the contour length of DNA in TPM. a) We tethered single DNA molecules coupled to beads to a surface and titrated HMGB1 (yellow) b) from 5nM-1000nM and found that HMGB1 reduces the height or <RMS> of the beads. We express the reduction in length as c) a change in <RMS> (nm) (blue), d) absolute reduction in DNA length (kbp) (red) or e) the fraction of DNA compacted (gray) by dividing the absolute reduction in DNA length (kbp) by the total length of DNA (kbp).

4.3.2 Divalent cation type alter the percentage of DNA compacted

HMGB1 has been shown to facilitate binding and enhance the catalytic activity of several DNA binding proteins in the presence of different divalent cations. The most abundant divalent cation in the nucleus is Mg^{2+}, but Ca^{2+} or Mn^{2+} are also added in biochemical assays to understand the function of the partner protein, but it is unclear how the divalent ions effect HMGB1 binding alone [8]. We tethered 539bp DNA beads and titrated HMGB1 from 5-225nM in the presence of 2.5mM Mg^{2+}, Ca^{2+}, or Mn^{2+}. We observed differential DNA compaction in the presence of particular divalents cations. The order of compaction based on divalent cation type was order as follows: $Mn^{2+} > Ca^{2+} > Mg^{2+}$ (Fig 4.2a-d). The divalent cation dependent variation in DNA compaction is probably the result of a change HMGB1's binding constant. Next we investigated how the length of DNA effects HMGB1-DNA compaction.

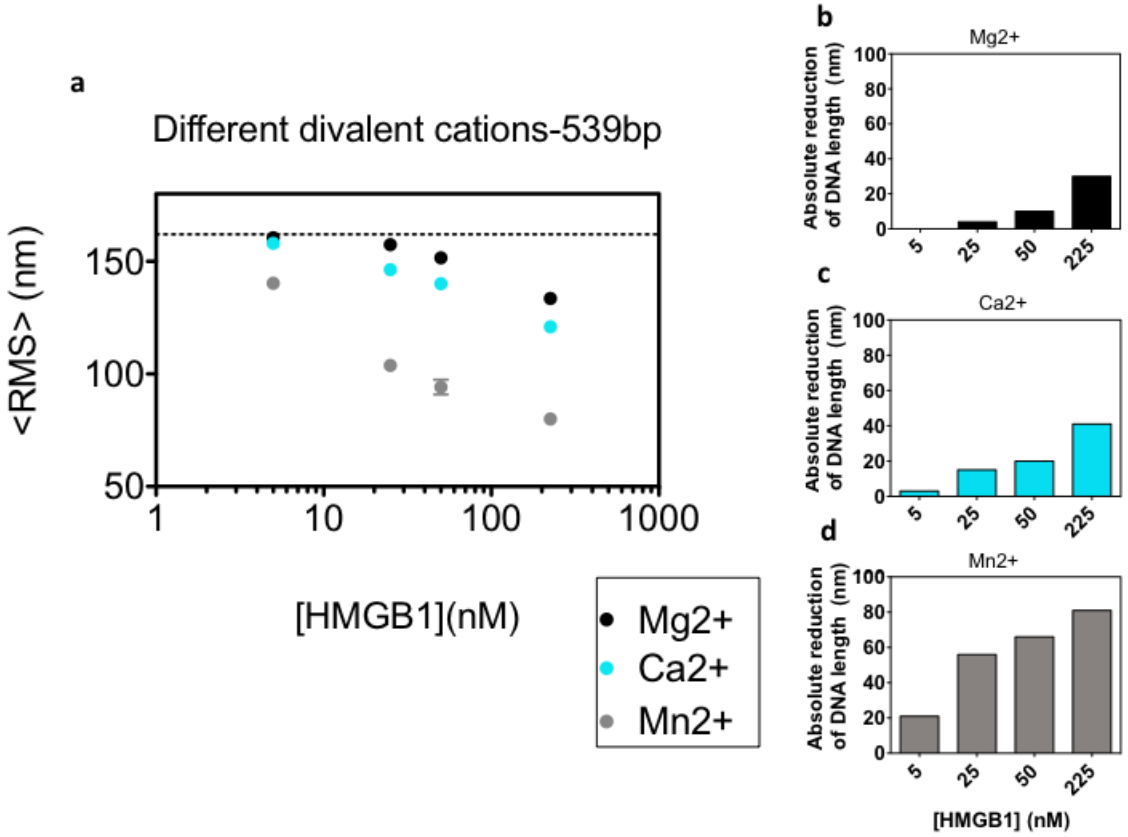

Figure 4.2: Divalent cation type differentially alters the magnitude of HMGB1-mediated DNA compaction. a) Titration of HMGB1 in the presence of Mg^{2+} (black), Ca^{2+} (light blue) or Mn^{2+} (gray). b-d) Absolute reduction in <RMS> in the presence of Mg^{2+}, Ca^{2+} or Mn^{2+}.

4.3.3 HMGB1 compacts the same percentage of DNA independent of contour length

To address the relationship between DNA length and HMGB1-DNA compaction, we assembled DNA tethered beads of eleven distinct lengths, 539, 736, 946, 1124, 1316, 1521, 1717, 1910, 2077, 2280, and 2900bps and titrated HMGB1 from 5nM-225nM HMGB1 (Fig 4.3a). Interestingly we find the percentage of DNA compaction at a particular HMGB1 concentration was conserved across all DNA lengths (Fig 4.3b).

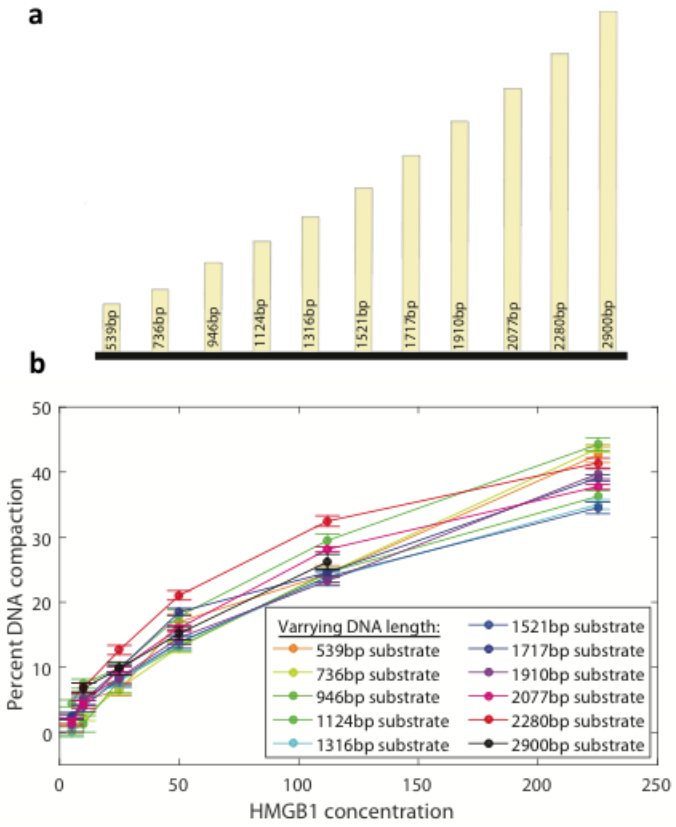

Figure 4.3: Fractional compaction by HMGB1 is conserved across DNA lengths. a) Schematic of DNA substrates varying in length from 539bp-2900bp. b) Percentage of DNA compacted.

4.4 Discussion

Our single molecule analysis shows that HMGB1 compacts DNA in the presence of distinct divalent cations and varying DNA lengths. We showed that HMGB1 compacts DNA in a concentration dependent manner, which confirms previous observations using applied force [9,10]. The compaction we observe does not provide insight into how each HMGB1 molecule contributes to the overall compaction, but previous studies have shown that HMGB1 melts dsDNA and reduces the end-to-end distance of a DNA segment called the persistence length within a DNA polymer [9]. The persistence length of DNA under thermal forces is 150bp or 50nm, and the addition of HMGB1 reduces the 50nm segment in a concentration dependent manner. We interpreted HMGB1-dependent DNA shortening as a result of the change in persistence length, but we directly observed DNA contour compaction. We also found that the amount of DNA compacted in Mg^{2+}, Ca^{2+}, or Mn^{2+} varied in magnitude for reasons that remain unclear. A probable mechanism is that the divalent cations change HMGB1's binding constant. One protein assembly facilitated by HMGB1 that requires divalent cations is the recombination activating genes RAG1 and RAG2, which assemble antigen-receptors in jawed-vertebrates by rearranging recombination signal sequences (RSSs). RAG protein biochemistry is Mg^{2+}-dependent, but most *in vitro* binding studies use Ca^{2+}, which prevents the RAG proteins from nicking RSSs and enhances the magnitude of DNA shifted in an electrophoretic mobility shift assay (EMSA) [8]. Our findings suggest the mechanism of Ca^{2+}-enhanced RAG and HMGB binding is a result of excess HMGB1 proteins forming RAG-HMGB1-RSS complexes and shifting more DNA in an EMSA than Mg^{2+}. Lastly, we observed that HMGB1 compact the same percentage of DNA independent of length. Future studies quantifying the number of HMGB1 molecules bound to each DNA molecule could provide further insight into the mean spacing of HMGB1 molecules along the DNA contour.

4.5 Methods

4.5.1 HMGB1 Purification

First, human HMGB1 was codon-optimized for bacterial expression by DNA2.0 and cloned into pJExpress vector. pJExpress-HMGB1 was transformed into BL21(DE3)pLysS cells

and plated on LB-agar with 50μg/mL kanamycin (kan). A single clone was picked for overnight liquid culture in 50mL LB-kan, and the following day 30mL of the overnight culture was diluted 100-fold in 3L of LB-kan. The cells were grown to an OD600 of 0.7 and induced by 1mM IPTG at 30 °C for 4h . The cells were harvested at 8,000rpm for 30 min and resuspended in 30mL of binding buffer (BB) (40mM Tris-HCl (pH 8.0), 500mM KCl, 0.25% Tween-20 and 50mM imididazole, 10μg/mL aprotinin, 10μg/mL leupeptin, 10μg/mL pepstatin A, 100μM PMSF and 5U/mL benzonase) for every 1L of cells. The resuspended cells were passed through a microfluidizer three consecutive times to lyse them. The cell lysate is centrifuged at 46,000g for 40 min, and the cleared lysate loaded onto a 5mL Ni-NTA column. The column was washed with 40mL of BB without benzonase followed by 20mL of wash buffer (40mM Tris-HCl pH 8.0, 65mM KCl, 50mM Imidazole), and eluted with 30mL elution buffer (40mM Tris-HCl (pH 8.0), 65mM KCl, 500mM Imidazole). Next, we used cation (SP) and anion (Q) exchange chromatography to eliminate truncated forms of HMGB1 using buffer solution A (Soln A: 40mM Tris-HCl (pH 8.0)) and solution B (Soln B: 40mM Tris-HCl (pH 8.0) and 1M KCl). The SP-column was equilibrated in 8% Soln B or 80mM KCl, and the Ni-NTA eluate was loaded onto the column, washed with 6mL of 8% Soln B and eluted in 1mL fractions with a gradient of 25mL Soln B from 8% to 65%. The fractions with HMGB1 were pooled and diluted to 80mM KCl, and loaded on a cation exchange Q-column equilibrated with 8% Soln B. The column was washed with 7mL of 8% Soln B, and eluted in 1mL fractions with a gradient of 40mL Soln B from 22% to 70%. The 1mL fractions with HMGB1 were pooled, concentrated, snap-frozen, and stored at $-80\,^{\circ}$C

4.5.2 DNAs

The labelled DNA substrates used in this study were amplified from the pUC19-NoRSS vector using the primers shown in Table S1. Eleven different DNA contour lengths of 539, 736, 946, 1124, 1316, 1521, 1717, 1910, 2077, 2280, and 2900bps were amplified with biotin and digoxigenin labelled primers. The labelled DNA was gel extracted, diluted to 200pM, and stored at $-20\,^{\circ}$C.

4.5.3 Tethered particle motion: Flow cell assembly, data acquisition and data analysis

TPM flow cell assembly begins by first drilling 12 holes into a pre-cleaned glass slide with a 1.5mm diamond tipped drill bit (CRLaurence) to create inputs and outputs for six channels. The drilled glass slide (Corning) and a partner coverslip (Fisher) are sonicated (Aquasonic-Model 75T) in 100% EtOH for 30 min, rinsed with ddH_2O, sonicated in 1M KOH for 30 min, rinsed with ddH_2O, sonicated in 5% HCl, rinsed with ddH_2O, and functionalized with epoxysilane (116mL Isopropanol (IPA), 1ml ddH_2O, 3ml of 3-glycidoxypropyldimethoxymethylsilane (Sigma), and 120 μL N,N-Dimethylbenzylamine (Sigma)) at RT for 1h and 30 min. The drilled glass slides and coverslips were washed in IPA, dried with N2 gas and baked at 100C for 30 min. Next, 0.02 in. ID/0.06 in. OD tygon microbore tubing (Cole-Parmer) is placed in each of the 12 holes of the drilled glass slide, designating an input and output for all six channels. Then five-minute epoxy (Devcon) is applied at the tubing-glass interface, and a piece of double-sided tape (0.12mm thick, Grace Bio-Labs) with six holes designating the channels of the flow cell is placed on the drilled glass slide and sandwiched between a coverslip. After the flow cell is assembled, polyclonal sheep anti-digoxigenin (Roche) is added to each of the six channels to a final concentration of 40ng/μL. The anti-digoxigenin (anti-dig) solution is incubated overnight at RT to allow covalent coupling of anti-dig to epoxy groups within the channels of the flow cell. Next, TPM assembly buffer (TPM-AB: 20mM Tris-Acetic Acid pH 8.0, 130mM KCl, 1mM MgCl2, 0.1mM DTT, 0.1mM EDTA, 20ug/ml Aceylated-BSA, and 3mg/ml Biotin-free Casein) is added to the flow cell's channels and incubated for 1h in order to flush out uncoupled anti-digoxigenin and quench active epoxy groups. Then biotin and digoxigenin labeled DNA is added to each channel to a final concentration of 5pg/μL and incubated for 20 min. The excess DNA is washed away with TPM-AB, then 0.49μm streptavidin coated polystyrene beads (Bangs Laboratories) diluted in TPM-AB to a final concentration 3 X 10^{10} beads/mL are added to each channel, and incubated for 10 min. The excess beads are washed away with TPM-AB and HMGB1 binding buffer (25mM Tris-HCl, 100mM KCl, 2mM MgCl2, 1mM DTT, 100μg/mL acetylated-BSA (A-BSA), 0.01% glycerol, 5% DMSO), completing the assembly of DNA-tethered beads.

The DNA tethered beads were imaged using brightfield microscopy (because the

polystyrene beads scatter light) on inverted Olympus IX71 microscopes with either a 100x oil objective, or a 60x oil objective with a 1.6x magnifier. A Basler A602f camera was used to acquire images at 60Hz; however, every other frame was dropped for a final frame rate of 30Hz (Note: This was implemented by a previous student to compare results imaged using a CCD camera with a frame rate of 30Hz. In these studies there is no particular reason we acquired at 30Hz). We track the DNA-tethered beads by cross-correlating each frame with the initial frame in a time series for each bead. This generates "raw" x and y positions of each bead's excursion relative to its anchor point. Drift was corrected from these raw data by subtracting the results of a low-pass first-order Butterworth filter with a cutoff frequency $f_{cB} = 0.05$ Hz. The root-mean-square (RMS) motion was obtained by applying a Gaussian filter with a -3 dB frequency of $f_{cG} = 0.0326$ Hz, corresponding to a 4 second or 120 frame standard deviation of the filter, to the mean-squared displacement of the data ($\overrightarrow{x}^2 + \overrightarrow{y}^2$). The RMS motion of the bead is the square root of the result of the convolution of ($\overrightarrow{x}^2 + \overrightarrow{y}^2$) and the Gaussian filter.

To define a "good" DNA tethered bead, we used symmetry-of-motion and length-of-motion checks used previously in [12]. We then track the good beads in buffer lacking HMGB1 for at least 350 sec, then stop tracking the beads and add a given amount of HMGB1 and continue tracking the same beads. This scheme allows for precision measurements of changes to the RMS motion of DNA-tethered beads by HMGB1. "Bad" beads which passed the initial checks are discarded if the beads stick to the surface, and if the bead's initial RMS is not within error to the previously measured RMS for a particular DNA length.

4.6 Appendix 1: TPM trajectories and distributions

HMGB1 trajectories and histograms for figures 4.1, 4.2 and 4.3 are shown below.

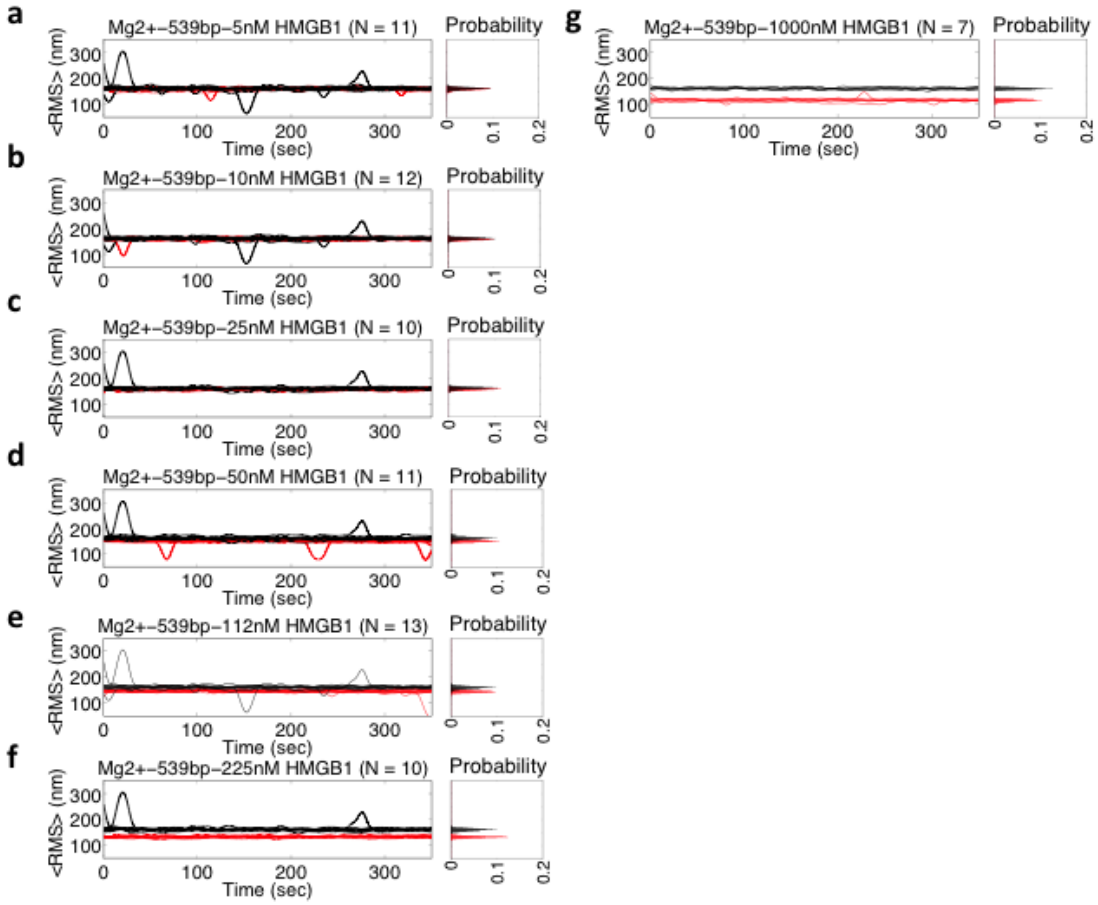

Figure 4.4: HMGB1 compacts DNA in a concentration dependent manner. HMGB1 was titrated over a concentration range of 5-1000nM in the presence of a 539bp DNA substrate. A-G are the trajectories that went into the data points in Fig. 4.1B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

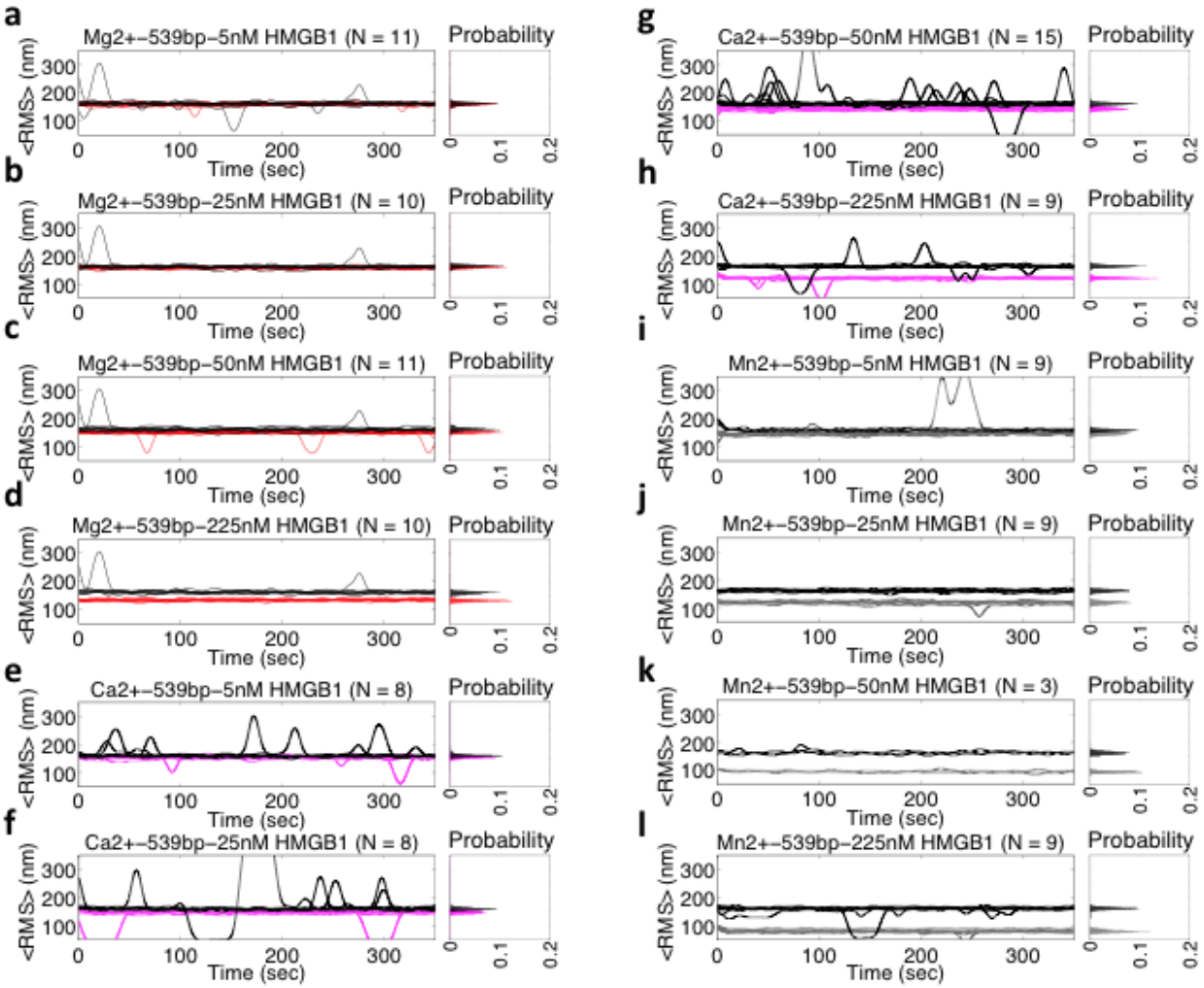

Figure 4.5: Divalent cation type differentially alters the magnitude of HMGB1-mediated DNA compaction. HMGB1 was titrated over a concentration range of 5-225nM in the presence of a 539bp DNA substrate, with Mg^{2+}, Ca^{2+} or Ca^{2+} in the reaction buffer. A-(i) are the trajectories that went into the data points in Fig. 4.2a. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red for Mg^{2+}, light blue for Ca^{2+} and dark gray for Mn^{2+}. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

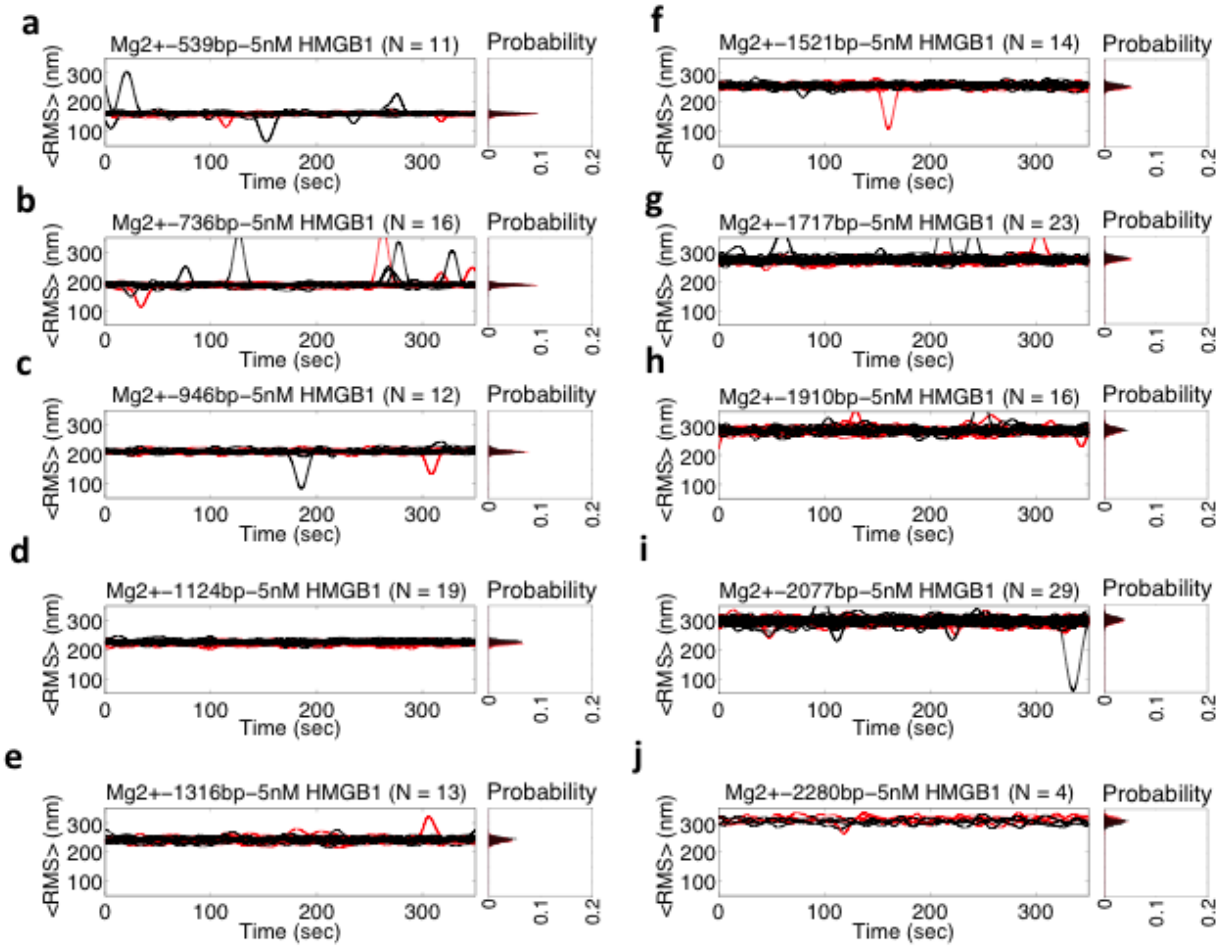

Figure 4.6: HMGB1 compacts the same fraction of DNA independent of contour length. 5nM of HMGB1 was added to DNA molecules ranging in length from 539bp-2280bp. A-(j) are the trajectories that went into the data points in Fig. 4.3B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

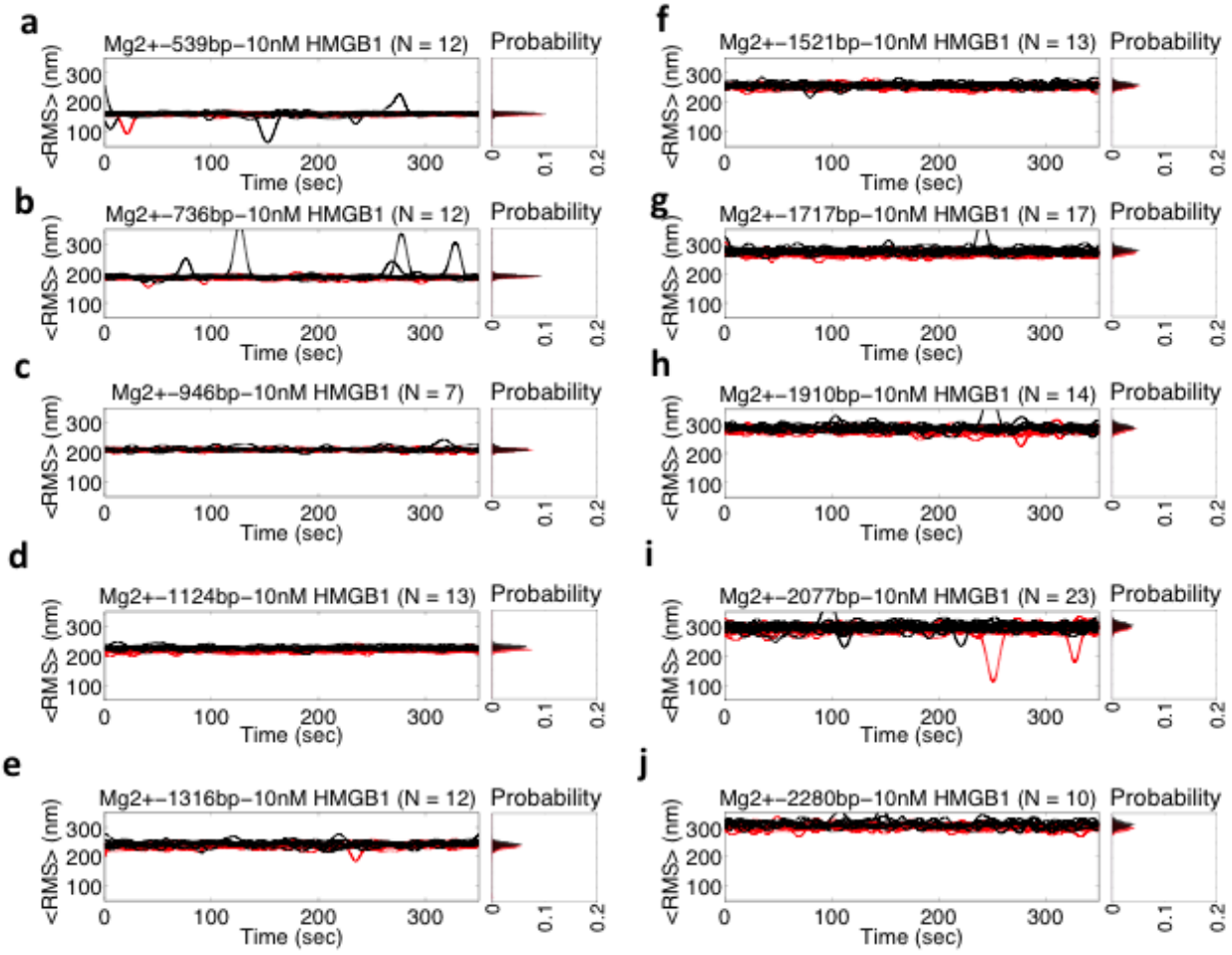

Figure 4.7: HMGB1 compacts the same fraction of DNA independent of contour length. 10nM of HMGB1 was added to DNA molecules ranging in length from 539bp-2280bp. A-(j) are the trajectories that went into the data points in Fig. 4.3B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajecto/ries before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

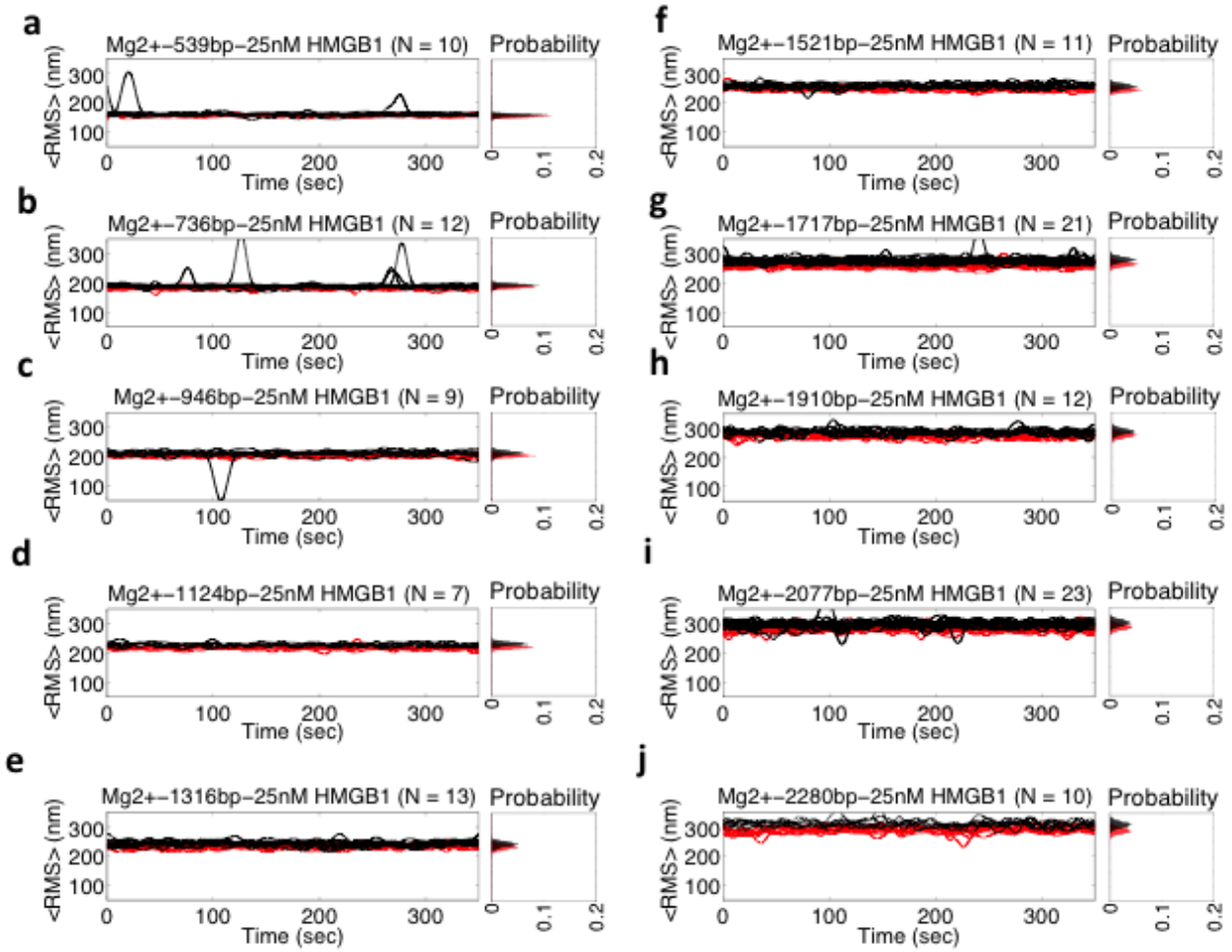

Figure 4.8: HMGB1 compacts the same fraction of DNA independent of contour length. 25nM of HMGB1 was added to DNA molecules ranging in length from 539bp-2280bp. A-(j) are the trajectories that went into the data points in Fig. 4.3B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

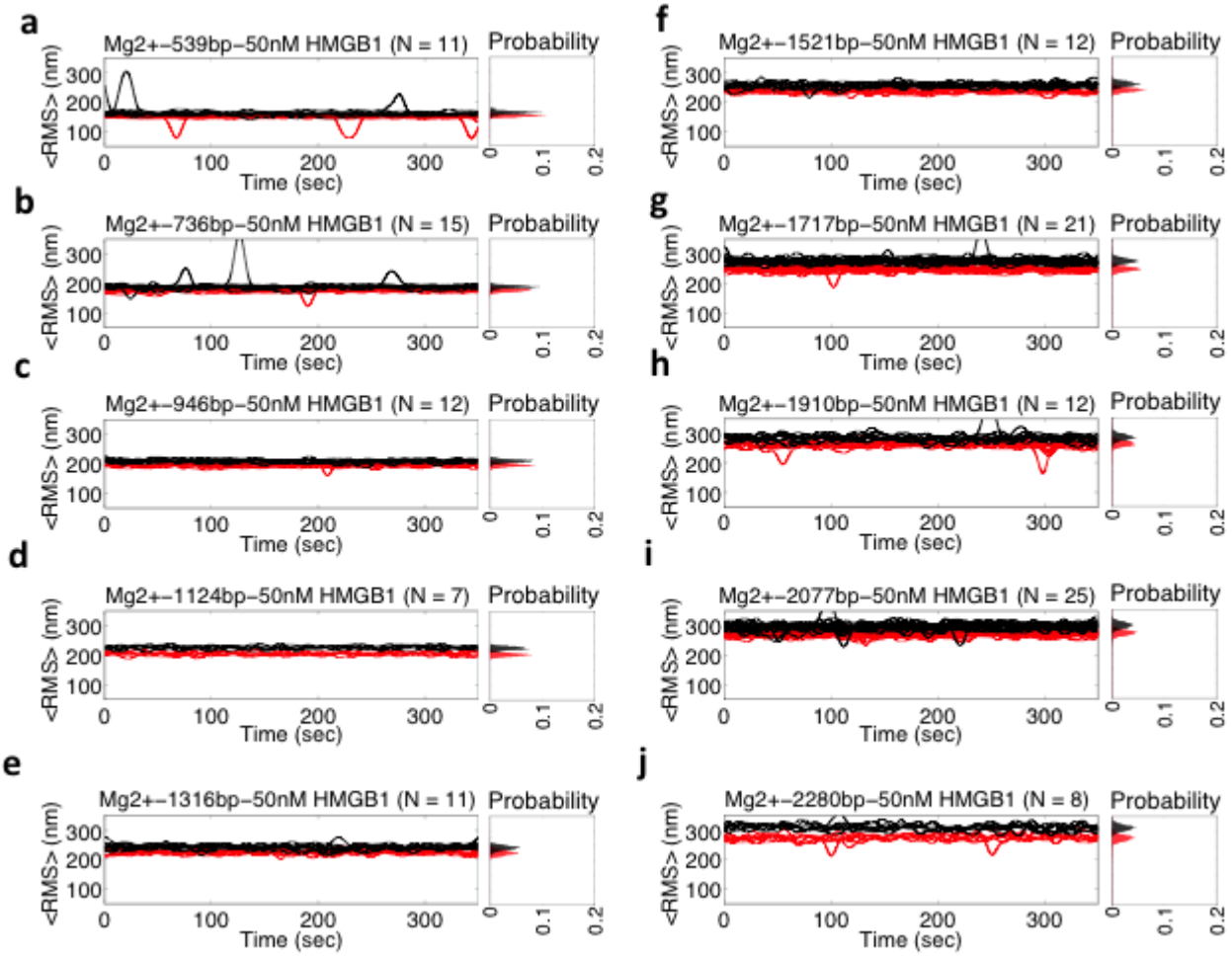

Figure 4.9: HMGB1 compacts the same fraction of DNA independent of contour length. 50nM of HMGB1 was added to DNA molecules ranging in length from 539bp-2280bp. A-(j) are the trajectories that went into the data points in Fig. 4.3B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

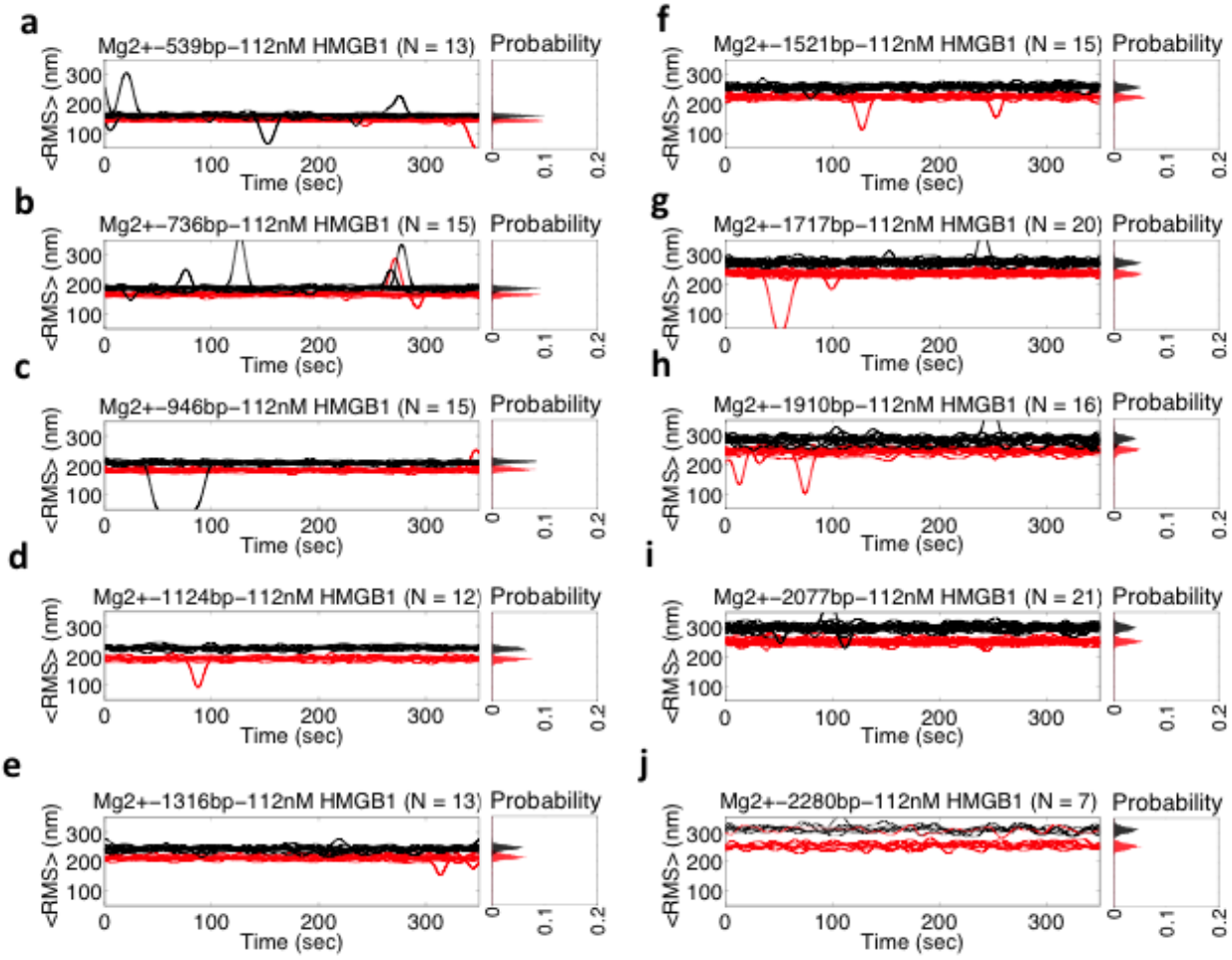

Figure 4.10: HMGB1 compacts the same fraction of DNA independent of contour length. 112nM of HMGB1 was added to DNA molecules ranging in length from 539bp-2280bp. A-(j) are the trajectories that went into the data points in Fig. 4.3B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

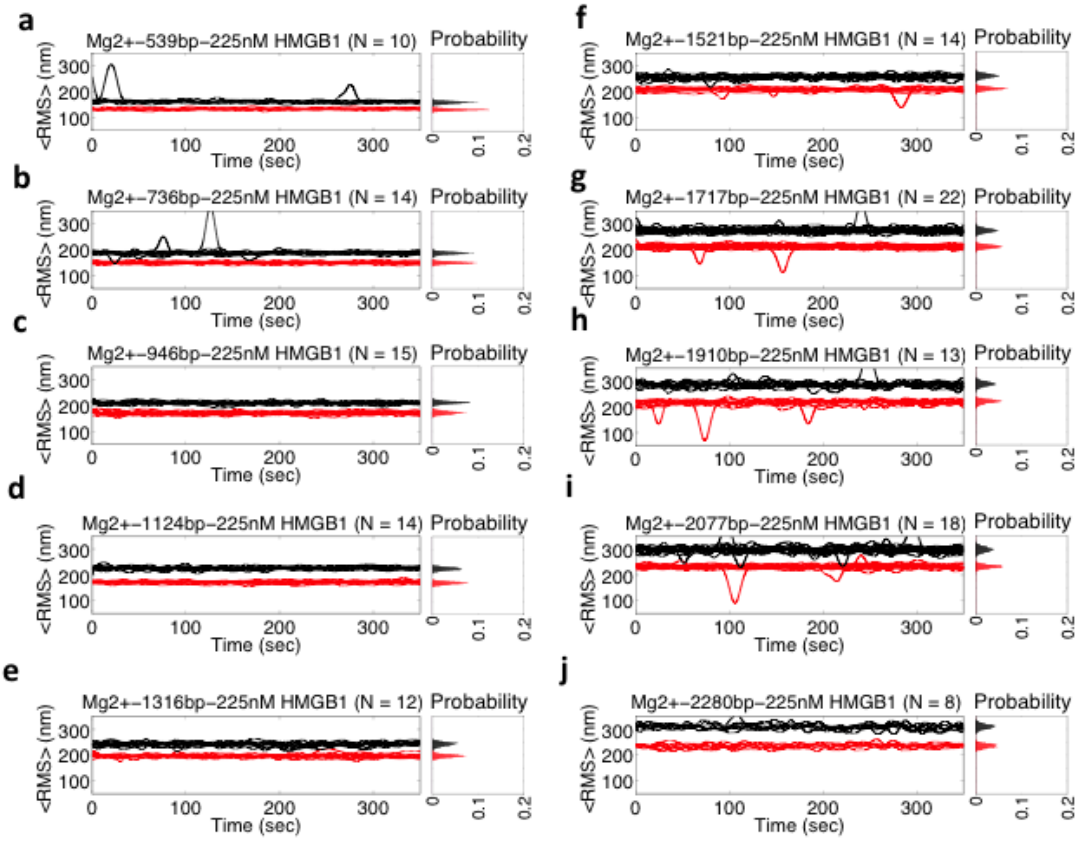

Figure 4.11: HMGB1 compacts the same fraction of DNA independent of contour length. 225nM of HMGB1 was added to DNA molecules ranging in length from 539bp-2280bp. A-(j) are the trajectories that went into the data points in Fig. 4.3B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before HMGB1 is added are black and the trajectories after HMGB1 is added are red. Spikes in the trajectories resulting in a decrease in the <RMS> are bead sticking events and an increase is another bead passing in the field.

4.7 References

[1]. Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility group chromosomal proteins. Mol. Cell. Biol. 19, 52375246.

[2] Johns, E. (ed.) (1982). HMG Chromosomal Proteins. pp. 1251, Academic Press, New

York.

[3] Bustin, M. (2001). Revised nomenclature for high mobility group (HMG) chromosomal proteins. Trends Biochem. Sci. 26, 152153.

[4] Bustin, M. and Reeves, R. (1996). High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. Prog. Nucleic Acid Res. Mol. Biol. 54, 35100.

[5] Laser, H., Bongards, C., Schuller, J., Heck, S., Johnsson, N. and Lehming, N. (2000). A new screen for protein interactions reveals that the Saccharomyces cerevisiae high mobility group proteins Nhp6A/B are involved in the regulation of the GAL1 promoter. Proc. Natl. Acad. Sci. U.S.A. 97, 1373213737.

[6] Jayaraman, L., Moorthy, N. C., Murthy, K. G., Manley, J. L., Bustin, M. and Prives, C. (1998). High mobility group protein-1 (HMG-1) is a unique activator of p53. Genes Dev. 12, 462472.

[7] Kruppa, M., Moir, R. D., Kolodrubetz, D. and Willis, I. M. (2001). Nhp6, an HMG1 protein, functions in SNR6 transcription by RNA polymerase III in S. cerevisiae. Mol. Cell, 7, 309318.

[8] D.C. van Gent, K. Hiom, T.T. Paull and M. Gellert Stimulation of V(D)J cleavage by High Mobility Group proteins EMBO J, 16 (1997), pp. 26652670

[9] McCauley,M.J., Zimmerman,J., Maher,L.J. III and Williams,M.C. (2007) HMGB binding to DNA: single and double box motifs.?J. Mol. Biol., 374, 9931004.

[10] Zhang, J., McCauley, M. J., Maher, L. J., III, Williams, M. C. and Israeloff, N. E. (2009). Mechanism of DNA flexibility enhancement by HMGB proteins. Nucleic Acids Res. 37, 11071114.

[11] Finzi,L. and Gelles,J. (1995) Measurement of Lactose repressor- mediated loop formation and breakdown in single DNA molecules. Science, 267, 378380.

[12] Han,L., Garcia,H.G., S., B., Towles,K.B., Beausang,J.F., Nelson,P.C. and Phillips,R. (2009) Concentration and length dependence of DNA looping in transcriptional regulation. PLoS One, 4, e5621.

Chapter 5

Distinct structural features of TFAM drive mitochondrial DNA packaging versus transcriptional activation

Work performed in collaboration with Huu Ngo and David Chan.

5.1 Abstract

TFAM is a DNA binding protein with multiple mitochondrial activities, including mitochondrial DNA (mtDNA) packaging and transcriptional activation at the light strand promoter (LSP) and heavy strand promoter 1 (HSP1). In crystallographic studies, TFAM has been shown to impose a U-turn on LSP DNA. It is unclear, however, whether this severe DNA distortion is relevant when TFAM is bound to either HSP1, where TFAM is proposed to bind in a reverse orientation, or to nonspecific DNA during genome packaging. We present crystal structures of TFAM bound to the heavy strand promoter 1 (HSP1) and to nonspecific DNA. In both, TFAM similarly distorts the DNA into a U-turn, and we experimentally demonstrate that TFAM binds to HSP1 in the opposite orientation from LSP. Transcription assays with engineered promoters indicate that this reverse orientation underlies the observation that transcription at the LSP promoters requires DNA bending, whereas transcription at HSP1 does not. Moreover, the crystal structures reveal TFAM dimerization. We show that TFAM dimerization requires binding to DNA, but is dispensable for DNA bending and transcriptional activation. However, TFAM dimerization plays an important role in DNA compaction, as assessed in a tethered particle motion assay that

measures DNA contour-length. Therefore, DNA bending is required for LSP transcription and mtDNA compaction but not for HSP1 transcription, and TFAM dimerization is required for mtDNA compaction but not transcription. We propose that TFAM dimerization enhances mtDNA compaction by promoting looping of mtDNA.

5.2 Introduction

TFAM (transcription factor A, mitochondrial) is a multi-functional DNA binding protein that is essential for transcriptional activation and mitochondrial DNA (mtDNA) organization [1-4]. As a sequence-specific transcription factor, TFAM plays a central role in production of transcripts from the light strand promoter (LSP) and the heavy strand promoter 1 (HSP1) of the mitochondrial genome. In *in vitro* transcription assays, normal levels of transcripts from LSP and HSP1 require TFAM [5, 6]. These mitochondrial transcripts encode for 13 essential components of the respiratory chain, and therefore TFAM is necessary for energy production from oxidative phosphorylation [1, 3]. In addition, TFAM is required for maintenance of mtDNA, because truncated transcripts from LSP are used to prime DNA synthesis during mtDNA replication. Consistent with these functions, mice lacking TFAM have impaired mtDNA transcription and inability to maintain mtDNA, resulting in bioenergetic failure and embryonic lethality [7].

We and others have previously solved the crystal structure of TFAM bound to LSP [7, 8]. The TFAM binding site at LSP is 22 base pairs long and is composed of two half-sites (Fig. 5.1A, B). TFAM contains two HMG (high mobility group)-box domains (HMG-box A and HMG-box B) that each intercalates into the minor groove of a half-site. Each intercalation contributes to distortion of the DNA, resulting in a dramatic U-turn of the LSP sequence (Fig. 5.1C). Between the two HMG-box domains is a helical linker with a positively charged surface that interacts with the negatively charged backbone of the DNA. The C-terminal tail of TFAM is required for activation of the transcriptional machinery. In the TFAM/LSP structure, the carboxyl terminal HMG-box B domain binds to the half-site distal from the transcriptional start site. Because the C-terminal tail is adjacent to HMG-box B, the U-turn in the LSP DNA enables the C-terminal tail to contact the transcriptional machinery. Consistent with this idea, TFAM mutants that are deficient in DNA bending are inactive for transcriptional activation at LSP [8]. In contrast, the same mutants are

fully active at HSP1. Based on sequence comparisons, TFAM has been suggested to bind to HSP1 in the reverse orientation compared to LSP [9, 10]. We proposed that in this reverse orientation, the C-terminal tail would be located near the half-site adjacent to the transcriptional machinery, rendering DNA bending unnecessary [8]. However, this model hinged on the expectation that the TFAM is reversed on HSP1 compared to LSP, an idea that lacks experimental evidence.

TFAM has an equally important function in organization of the mitochondrial genome. Located in the mitochondrial matrix and lacking histone-related proteins, mtDNA is organized into compact mtDNA/protein structures termed nucleoids. TFAM is one of the most abundant proteins of the nucleoid [11, 12] and is thought to coat the entire mitochondrial genome and compact it [13, 14]. Based on the TFAM/LSP structure, it is tempting to assume that TFAM also imparts a U-turn when bound nonspecifically to mtDNA. In this manner, coating of the mitochondrial genome by TFAM would shorten its apparent length by imposing regular sharp bends. The structurally unrelated HU family of proteins also imparts U-turns onto DNA and may play an analogous role in the architecture of procaryotic genomes [15, 16]. However, it is unclear whether the structure of TFAM bound to LSP, where it binds in a sequence-specific mode, serves as a good model for TFAM coating the mitochondrial genome, where it functions as a nonspecific DNA binding protein. Although TFAM bends DNA at LSP, it has been suggested that such rigid DNA bends are not applicable when TFAM is bound to nonspecific DNA [17]. Moreover, it remains to be determined whether DNA bending is a sufficient mechanism to account for the dramatic condensation of mtDNA within a nucleoid compared to naked circular DNA.

In this study, we present additional crystal structures of TFAM bound to HSP1 and to nonspecific mtDNA, and show that in both cases TFAM imposes a U-turn on the DNA. We experimentally confirm that TFAM binds HSP1 in the opposite orientation compared to the TFAM/LSP structure and provide evidence that this orientation difference underlies the different requirements of the two promoters for DNA bending. In the four existing TFAM structures, we observe dimerization of TFAM via the same protein-protein interface. We show that dimerization is not needed for DNA bending or transcriptional activation, but is necessary for full compaction of DNA. These results suggest that TFAM dimerization provides an additional compaction mechanism beyond DNA bending to help organize mtDNA into nucleoids.

5.3 Results

5.3.1 Comparison of TFAM bound to specific and nonspecific DNA

We previously solved the crystal structure of TFAM bound to its cognate binding site on LSP [8] (Fig. 5.1A-C). This structure indicates that TFAM imposes a U-turn on LSP DNA upon binding, but the generality of this observation is unknown. It is unclear whether the TFAM/LSP structure is representative of how TFAM binds to HSP1, where DNA bending by TFAM is dispensable for transcriptional activation [8]. Even more important, it is unclear whether TFAM also imparts a U-turn when it binds nonspecific mtDNA, where it is thought to play an important role in DNA packaging. To address these issues, we generated crystal structures of TFAM bound to HSP1 mtDNA and non-specific mtDNA.

Initial attempt to generate such structures yielded small and poorly diffracting crystals. Because the TFAM/LSP structure indicates that the last 9 residues of TFAM are disordered [7,8], we repeated crystallization trials with a new TFAM construct (residues 43-237) lacking these disordered residues. We obtained well-diffracting crystals of TFAM bound to a 22 bp double- stranded DNA fragment from HSP1 (Fig. 5.1D; termed the TFAM/HSP1 structure). This DNA fragment contains the TFAM binding site determined by DNAse footprinting [18]. In addition, we crystallized TFAM in complex, with a 22 bp fragment from the ATPase6 gene as an example of TFAM binding nonspecifically to mtDNA (Fig. 5.1E; termed the TFAM/nonspecific DNA structure). Both the TFAM/HSP1 and TFAM/nonspecific DNA structures were solved by molecular replacement, using the coordinates of TFAM/LSP [19] as the search model. The structures of TFAM/HSP1 and TFAM/nonspecific DNA were solved at a resolution of 2.9 $\mathring{A}$ and 2.8 $\mathring{A}$, respectively. For both, the final electron density maps were of sufficient quality to visualize all the protein residues and the entire DNA fragment. Model building and refinement led to final Rwork and Rfree values of 22.87% and 26.61% for TFAM/HSP1, and 21.37% and 26.11% for TFAM/nonspecific DNA. Details of the crystallographic analyses are presented in Table 1.

The TFAM/HSP1 and TFAM/nonspecific DNA structures are highly similar to TFAM/LSP [8, 19]. Superimposition against the TFAM/LSP structure generates RMSD values of 1.056 $\mathring{A}$ for TFAM/HSP1 and 0.951 $\mathring{A}$ for TFAM/nonspecific DNA (Fig. 5.1F). In each case, TFAM binding imposes a full U-turn on the DNA. This U-turn is caused by the wedging of the two HMG-box domains of TFAM into the minor groove of the DNA.

Each HMG-box domain consists of three helices that are arranged in an L-shaped configuration stabilized by hydrophobic interactions (Fig. 5.1C-E). The concave surface of each HMG-box domain intercalates into the DNA minor groove, with contributions from Leu58 and Ile81 of HMG-box A and Asn163, Pro178, and Leu182 of HMG-box B (Fig. 5.1C-E). Analysis of DNA parameters indicates that distortion of the DNA double helix is maximal at the bases contacted by Leu58 and Leu182 (Fig. 5.1G). A helical linker with a positively charged surface connects the two HMG-box domains and packs against the overlaying DNA [8, 19]. As originally described for TFAM/LSP, the two intercalating HMG- box domains and the linker, acting as a pivot, work in concert to impart a U-turn on the bound DNA.

5.3.2 TFAM imposes a U-turn regardless of mtDNA sequence

Taken together, the set of crystal structures indicate that TFAM bends DNA into a U-turn, whether bound specifically to LSP or HSP1, or bound nonspecifically to a random mtDNA fragment. To determine whether this DNA distortion also occurs in solution, we used an established FRET-based assay to measure DNA bending [8]. To construct the FRET sensor, Cy3 (donor) and Cy5 (acceptor) fluorophores were covalently attached to opposite ends of the LSP, HSP1, and ATPase6 (nonspecific) DNA fragments. Addition of TFAM to the labeled, double- stranded DNA resulted in a dose-dependent increase in acceptor emission and a decrease in donor emission. Control experiments confirmed that the acceptor emission depended on the presence of both the donor fluorophore and TFAM. For each DNA fragment, the maximal FRET efficiency corresponded to a distance of 59 $\AA$ between the DNA ends, in agreement with the crystal structures of the complexes (Fig. 5.1F, H). These data indicate that all three DNA templates are bent to a similar degree by TFAM, suggesting that the formation of U-shaped DNA is relevant not only in transcriptional activation, but also in mtDNA packaging.

5.3.3 Reverse binding orientations of TFAM underlies distinct dependence of LSP and HSP1 to DNA bending

Based on DNA sequence analysis of the promoters, TFAM has been proposed to bind in opposite orientations on LSP versus HSP1 [9, 10]. This idea has implications about the mechanism of transcriptional activation by TFAM, and therefore it is important to obtain

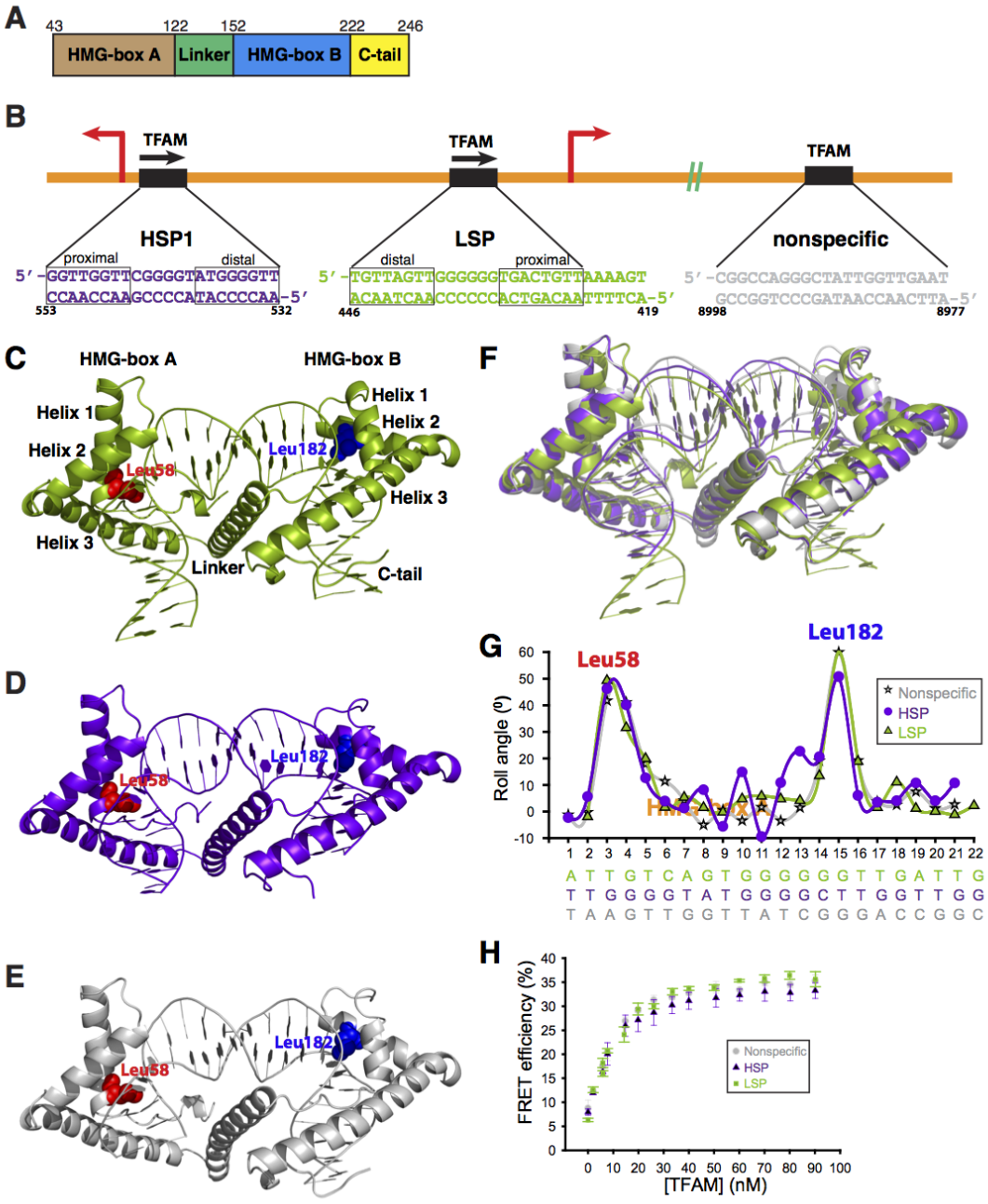

Figure 5.1: **Overview of the TFAM-mtDNA complexes** (A) The domain structure of mature TFAM. Residues 1-42 constitute the mitochondrial targeting sequence that is cleaved upon import of TFAM into the mitochondrial matrix. (B) Schematic of DNA sequences bound within TFAM crystals. Note the different orientations of TFAM on LSP versus HSP1. The nonspecific sequence is from the ATPase6 gene. (C), (A), (E) Side view of the TFAM/LSP, TFAM/HSP1, and TFAM/nonspecific DNA complexes, respectively. The major intercalating residues, Leu58 and Leu182, are highlighted. The DNA fragments are color-coded as in (B). (F) Superimposition of TFAM crystal structures, color-coded as in (B). (G) Comparison of roll angle values for TFAM/LSP, TFAM/HSP1, and TFAM/nonspecific DNA. (H) FRET assay for DNA bending by wild-type TFAM using three different DNA templates: LSP, HSP1, and nonspecific DNA. Data points are the average of three independent experiments, with error bars representing standard deviations.

direct experimental support. To unambiguously assign the orientation of the DNA sequence in the TFAM/HSP1 structure, we substituted thymine at position 550 with bromo-uracil (Fig. 5.2A, B). This substitution uniquely labels HSP1 DNA at the half-site proximal to the transcriptional initiation site. Bromine has significant anomalous scattering signal, and X-ray data from isomorphous crystals showed an anomalous peak (>5σ) at the expected position for bromine adjacent to the HMG-box B domain (Fig. 5.2C, D). These data confirm that TFAM binds HSP1 with HMG-box B occupying the proximal half-site, which is the opposite orientation of that found in the TFAM/LSP complex.

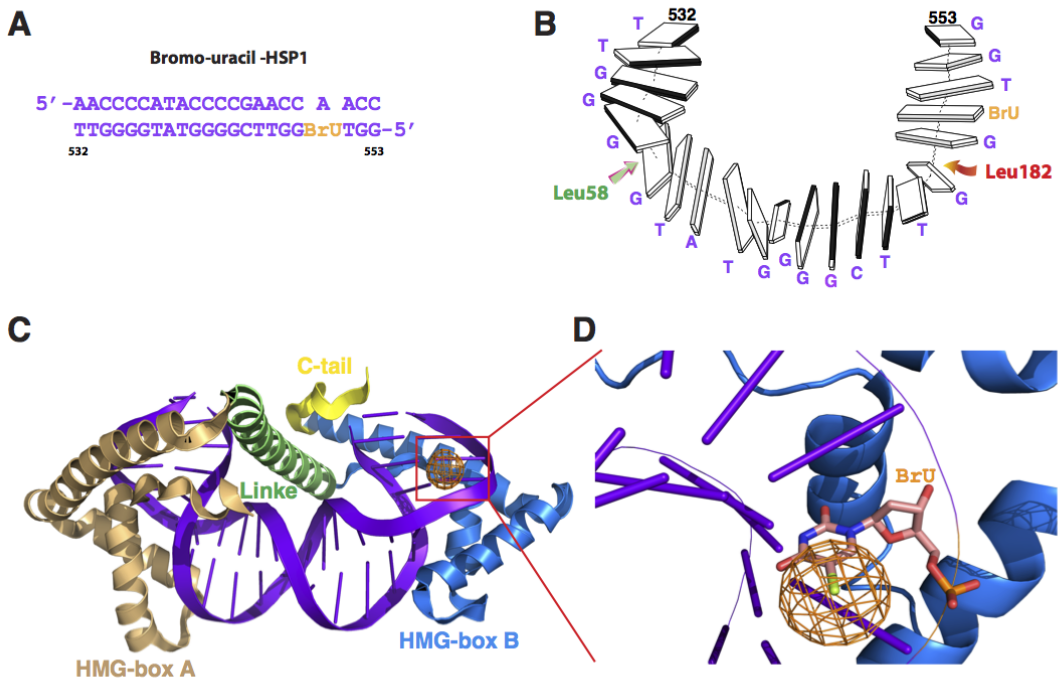

Figure 5.2: **TFAM binds HSP1 in a reverse orientation** (A) Thymine 550 within HSP1 was replaced with bromo-uracil to label the proximal TFAM half-site. (B) A presentation of the 22 bp HSP1 structure, showing its U-turn shape and the location of intercalating residues. (C) The TFAM/HSP1 complex showing location of the anomalous signal of bromo-uracil (orange) relative to the domains of TFAM. The anomalous signal is adjacent to HMG-box B. (D) Close-up view of the anomalous electronic density in a Friedel-pair difference map revealing a $> 5\sigma$ peak (orange).

We used *in vitro* transcription reactions with LSP or HSP1 templates to test the idea that TFAM has different requirement on these promoters. Transcription from the LSP and HSP1 promoters can be reconstituted *in vitro* by providing the DNA templates

with mitochondrial RNA polymerase (mtRNAP), TFB2M, and TFAM [5]. We previously generated a mutant (L6) that contained point mutations in the helical linker connecting the two HMG-box domains of TFAM [8]. This mutant showed the most severe DNA bending defect from a panel of mutants. Like other DNA bending mutants, L6 is defective in activating transcription from LSP. However, this mutant is able to activate transcription at HSP1 as effectively as wild-type TFAM [8] (Fig. 5.3B, C).

To probe the basis for this defect, we asked whether HSP1 could be converted into a promoter that is sensitive to DNA bending by TFAM. In the first approach, we reversed the TFAM binding site at HSP1, so that HMG-box B would now bind at the distal half-site (Fig. 5.3A, EP1 template). In the second approach, we replaced the TFAM binding site in HSP1 with the corresponding region from LSP, again forcing TFAM to bind in a reverse orientation compared to native HSP1 (Fig. 5.3A, EP2 template). When tested with *in vitro* transcription assays, both these modified HSP1 promoters are activated by wild-type TFAM. However, the bending defective L6 mutant shows a several-fold decrease in transcriptional activation (Fig. 5.3B, C).

5.3.4 Dimerization of TFAM

Along with the two new TFAM structures reported here, there are now four structures of TFAM bound to DNA. Remarkably, all the structures reveal dimerization of TFAM stabilized by an identical dimer interface. In each case, the dimer interface occurs between HMG-box A of two separate TFAM molecules. HMG-box A consists of three helices arranged in an L-shape, with helix 3 forming the long edge. Fig. 5.4A shows the interface found within the asymmetric unit of TFAM/HSP1. The convex surfaces of two HMG-box A domains come into contact, with helix 3 from each domain paired in an antiparallel manner (Fig. 5.4A). The antiparallel interface shows a series of polar and electrostatic interactions mainly mediated by residues Lys95, Tyr99, Glu106, Glu112, and Arg116 of helix 3 (Fig. 5.4A, B). In the helix 3/helix 3 interface, Lys95 forms a salt bridge with Glu112, and Tyr99 forms a hydrogen bond with Arg116. Between helices 3 and 1, Arg116 interacts with Glu63, and Glu106 interacts with Arg59. With over 30 residues involved, the dimer interface has a total buried surface area of approximately 1179 $\AA 2$ (Fig. 5.4C). Crystal packing analysis of the TFAM/LSP 22 bp [19], TFAM/LSP 28 bp [8], and TFAM/nonspecific DNA structures reveals identical, antiparallel helical interfaces that superimpose well onto the

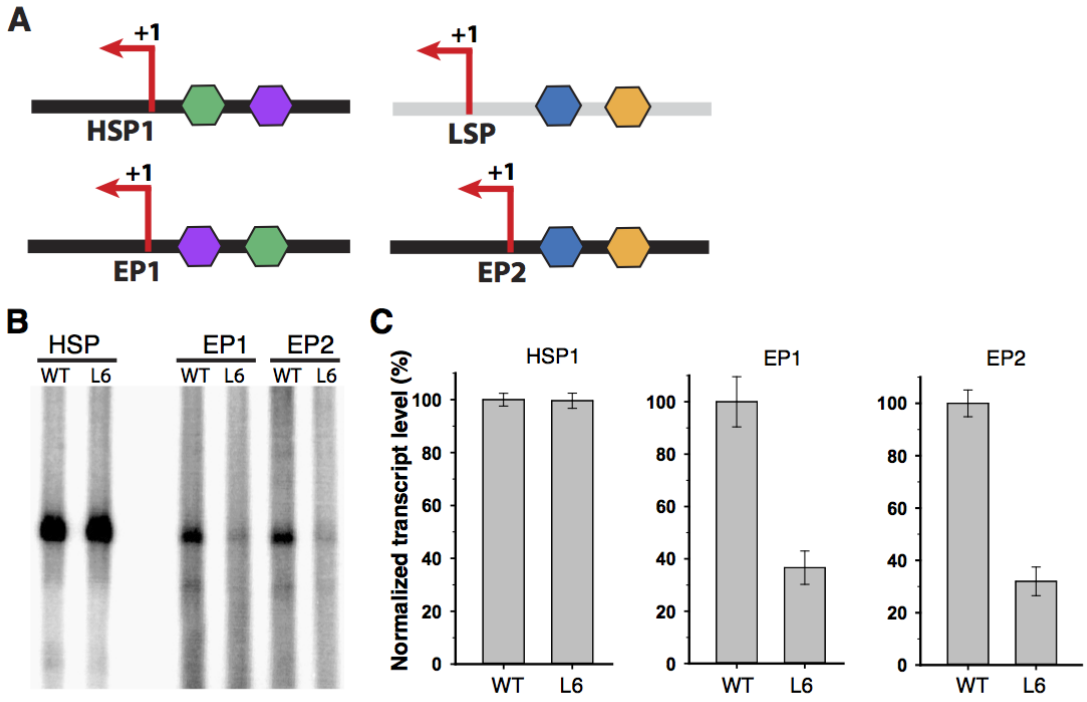

Figure 5.3: **Conversion of HSP1 into a promoter dependent on DNA bending**(A) Schematic of HSP1, LSP, and two engineered promoters (EP1, and EP2) derived from HSP1. EP1 and EP2 were designed to reverse the orientation of TFAM on HSP1. (B) Representative transcription reactions with wildtype TFAM and the L6 mutant that is deficient in DNA bending. (C) Quantification of transcription reactions, with error bars representing standard deviations from three independent experiments. Values are normalized to that of wildtype TFAM.

interface found in TFAM/HSP1 (Fig. 5.4D). This interface is unrelated to the C-terminal region previously proposed as a candidate dimerization motif [14].

5.3.5 TFAM Dimerization is not required for DNA bending or transcriptional activation

To test the physiological function of dimerization, we generated a TFAM mutant with five substitutions (K95A, Y99F, E106A, E112A, and R116A; hereafter termed "dimer mutant") designed to disrupt polar and electrostatic interactions at the interface. To assess TFAM dimerization, we developed a FRET-based assay to measure TFAM/TFAM contact. We covalently labeled TFAM molecules with either a donor fluorophore (Alexa Fluor 488) or an acceptor fluorophore (Alexa Fluor 595) using cysteine-maleimide chemistry at the single cysteine residue at position 49, which is close to the dimer interface. In the absence of DNA, no FRET was detected between the two labeled populations (red trace, Fig. 5.5A). However, in the presence of DNA, we found a decrease in donor emission and an increase in acceptor emission (magenta trace, Fig. 5.5A). This FRET signal could be competed off by excess unlabeled TFAM (Fig. 5.6). This FRET signal was abolished in the dimer mutant (blue trace, Fig. 5.5A), suggesting loss of dimerization. Interestingly, the dimerization of wild-type TFAM was not only dependent on DNA, but also on the length of the DNA. Testing a range of DNA lengths (100, 150, 200, 300, to 400 bp), we found robust TFAM dimerization only with DNA fragments $\geq$ 200 base pairs (Fig. 5.5B).

Although defective in dimerization, the dimer mutant is fully functional for DNA binding and bending on LSP, HSP1, or nonspecific DNA templates (Fig. 5.5C). Moreover, the dimer mutant is as efficient as wild-type TFAM in transcriptional activation from either the LSP or HSP1 template (Fig. 5.5D, E). Taken together, these data suggest that dimerization is dispensable for both DNA bending and transcriptional activation.

5.3.6 DNA bending and TFAM dimerization are required for efficient DNA compaction

In addition to transcriptional activation, the other major function of TFAM is mtDNA compaction. In the latter function, TFAM is thought to interact with mtDNA in a global, nonspecific manner. To analyze the ability of TFAM mutants to compact DNA, we utilized

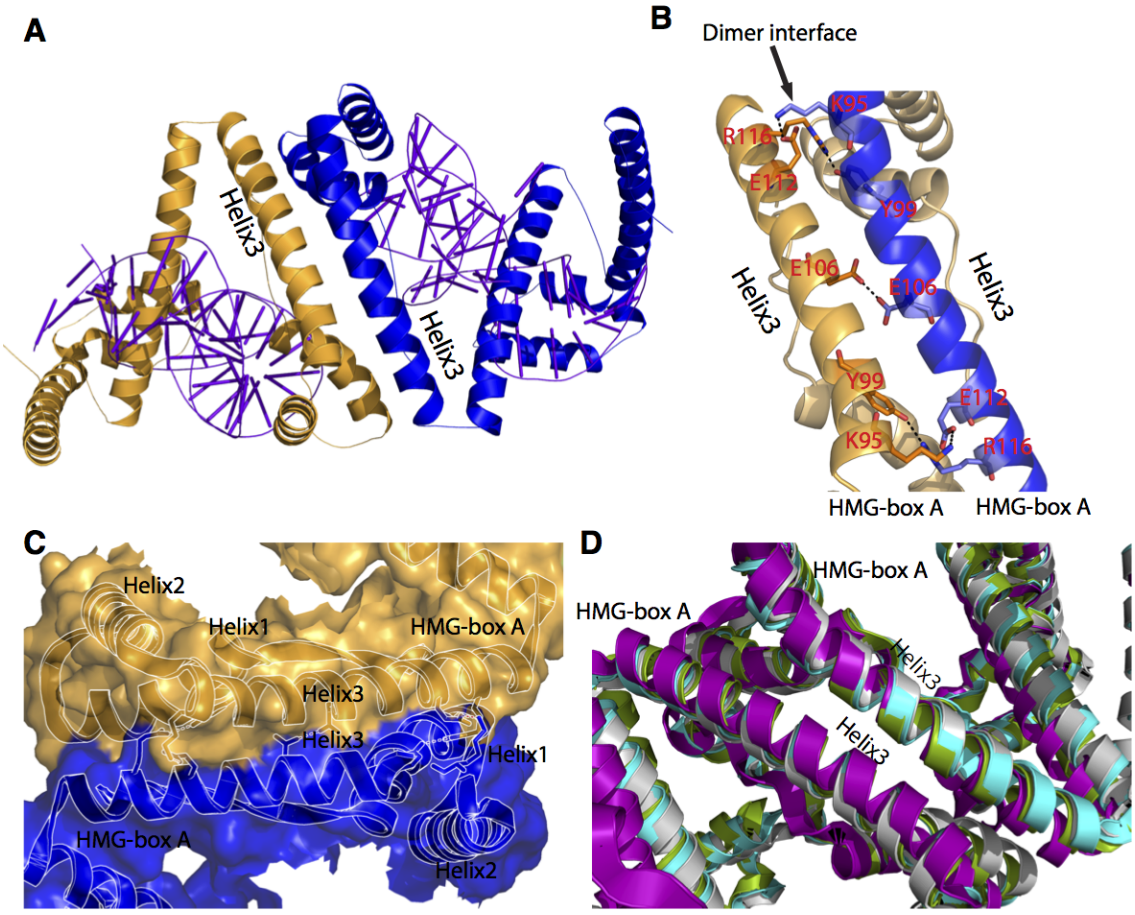

Figure 5.4: **Dimerization interface** (A) Overview of two molecules of TFAM forming a dimer in the TFAM/HSP1crystal structure. Each TFAM molecule is bound to its own DNA fragment. Helix 3 from one HMG-box A domain forms an antiparallel interface with the corresponding helix 3 from another molecule. (B) Close-up of the antiparallel dimerization interface. Residues involved in hydrogen bonds and salt bridges are labeled. (C) Space-filling model of the dimerization interface. (D) Superimposition of the dimerization interfaces from all four TFAM-DNA structure: TFAM/LSP-28 bp (green, pdb:3TMM), TFAM/LSP-22 bp (cyan, pdb:3TQ6), TFAM/HSP1-22 bp (purple), and TFAM/nonspecific DNA-22 bp (grey). R.M.S.D. values relative to TFAM/LSP-22bp are as follows: TFAM/LSP-28 bp (0.887), TFAM/HSP1-22 bp (1.056) and TFAM/nonspecific DNA- 22 bp (0.951).

a tethered particle motion (TPM) assay (Fig. 5.7A), in which the contour lengths of single DNA molecules are measured upon interaction with TFAM [17]. In this assay, one end of a DNA molecule is attached to a bead, and the other end is immobilized onto a slide. Because the DNA molecule has a defined length, the tethered bead will have a characteristic radius of motion that can be quantified as a root-mean-square (RMS) value (see Appendix 4.7 for more details). Addition of TFAM to the system reduces the radius of motion by decreasing the end-to-end length of the DNA tether, a reflection of DNA compaction (Fig. 5.7A). Using a 1910 base pair DNA tether, we found progressively lower RMS values with increasing TFAM concentrations (Fig. 5.7B). The naked DNA tether had an RMS value of 295 nm; this value was reduced to 164 nm under saturating TFAM concentrations (1 μ M). By testing a panel of DNA lengths ranging from 539 bp to 1910 bp, we found that TFAM at saturating concentration reduces the apparent lengths proportionally to the original length of the tether (Fig. 5.7C). To compare these DNA compaction effects, we generated a DNA calibration curve for naked DNA (Fig. 5.8A). This calibration curve allowed us to convert RMS values in the presence of TFAM to apparent DNA length, and thereby calculate the extent of compaction in terms of base pairs (Fig. 8B). Regardless of the original DNA length, TFAM shortened the DNA contour-length by 64-73% (Fig. 5.7D)

This level of DNA compaction exceeds the maximal shortening that can be accounted for by DNA bending, and prompted us to examine the structural basis for DNA compaction in the TPM assay. In the TFAM crystal structures, a 22 bp DNA fragment is shortened by 50% due to formation of a U-turn. Multiple regions of TFAM are important for bending mtDNA [8, 19], including both HMG-box domains and the intervening helical linker. Using the 1910 bp DNA tether, we compared the compaction abilities of TFAM mutants previously established as having a DNA bending defect [8]. While wild-type TFAM compacted DNA molecules to 164 nm, TFAM truncation mutants containing either HMG-box A alone or HMG-box B alone shortened DNA to only 212 nm and 227 nm, respectively (Fig. 5.7E). The linker region of TFAM coordinates the spacing and orientation of the two HMG-box domains and is important for DNA bending. The L6 linker mutant also shows a defect in DNA compaction. Hence, both HMG-box domains, as well as the intervening linker are necessary to achieve efficient DNA compaction. This result is not surprising, given that an important aspect of DNA compaction should be the formation of DNA U-turns.

To examine whether TFAM dimerization might affect DNA compaction, we tested

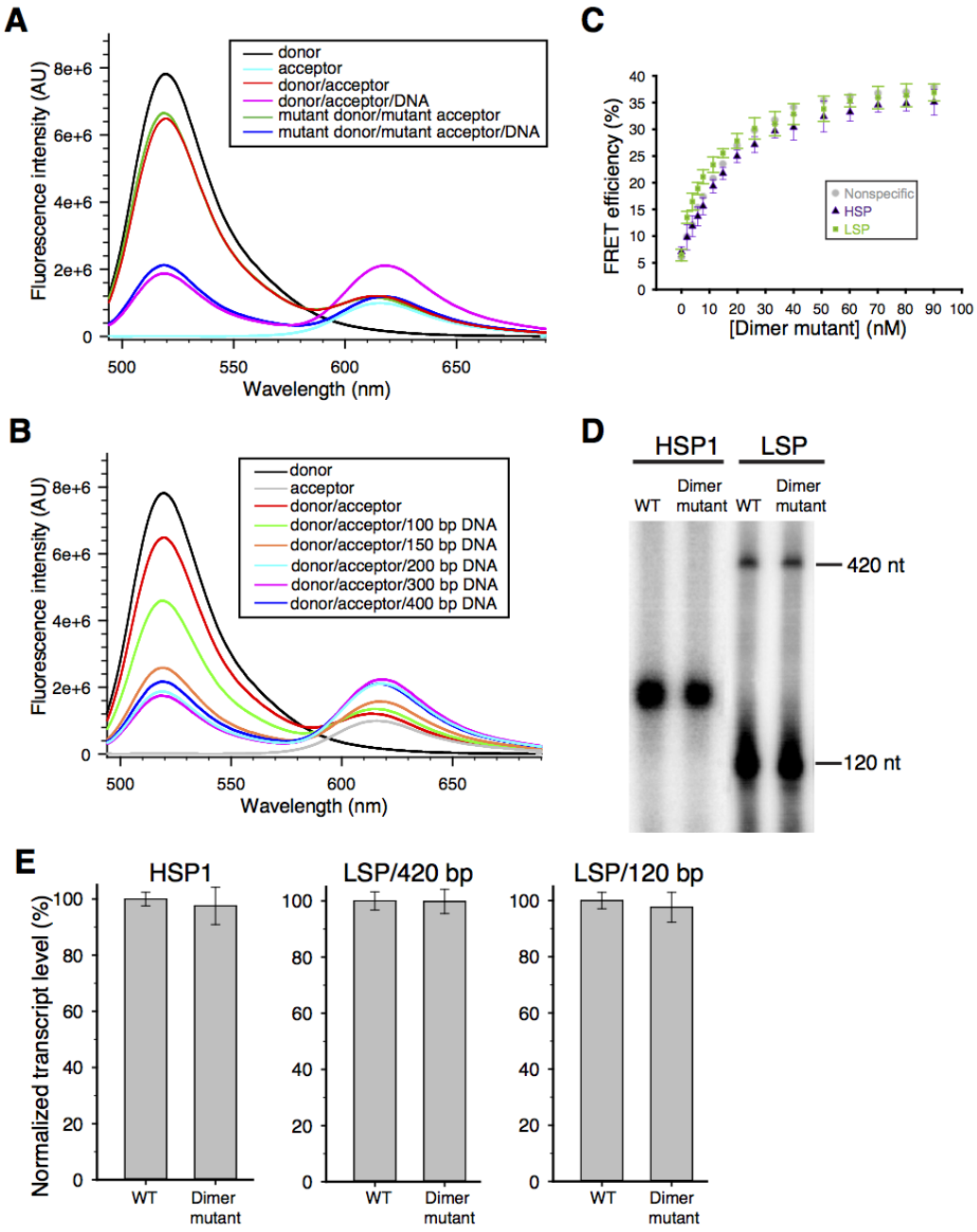

Figure 5.5: **Biochemical analysis of TFAM dimerization** (A) Emission spectra in a FRET assay measuring the physical interaction between TFAM molecules. Reactions contained Alexa Fluor 488 (donor)-labeled and/or Alexa Fluor 594 (acceptor)-labeled TFAM. Fluorescence emission spectra shows FRET signal only in the presence of plasmid DNA (pink trace). This signal is abolished in the dimer mutant (purple trace). (B) Emission spectra of wild-type TFAM incubated with linear DNA of varying lengths. (C) DNA bending by the dimer mutant on three templates. Data points are the average of three independent experiments, with error bars representing standard deviations. (D) Representative transcription assay using wild-type TFAM or the dimer mutant. The LSP template generates a 420 nt full-length (run-off) transcript and a truncated 120 nt transcript. (E) Quantification of transcription reactions with error bars representing standard deviations from three independent experiments.

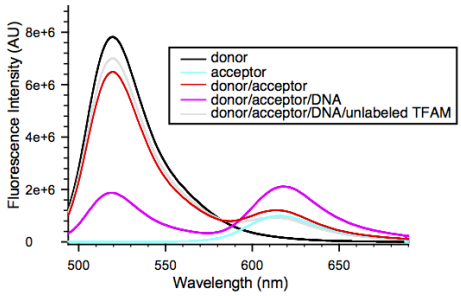

Figure 5.6: **The FRET signal from TFAM dimerization can be competed out by excess unlabeled TFAM** In the chase experiment (gray line), the signal between 400 nM TFAM-Alexa Fluor 488 (donor) and 100 nM TFAM-Alexa Fluor 594 (acceptor) was competed by 5 μ M unlabeled TFAM

the dimer mutant in the TPM assay. We found that the dimer mutant has a substantial defect in DNA compaction, with an RMS value of 193 nm (Fig. 5.7E). This DNA compaction defect is not due to loss of DNA binding, because the mutant binds to DNA with the same affinity as wild-type TFAM (Kd 7 nM) and has no defect in either DNA bending or transcriptional activation (Fig. 5.5C-E). These results therefore uncover TFAM dimerization as an important aspect of DNA compaction.

5.4 Discussion

Our study provides several important insights into the structural biology of TFAM. First, by providing additional TFAM structures, we show that TFAM imposes a U-turn on mtDNA, whether it is bound at LSP, HSP1, or nonspecific DNA. Therefore, extreme DNA bending is a constitutive feature of TFAM binding. It has been suggested that DNA U-turns may not be relevant when TFAM is bound to nonspecific DNA [17]. However, we show that TFAM mutants defective in bending have a reduced ability to compact DNA. Second, we provide evidence that the different orientations of the TFAM binding sites at LSP compared to HSP1 underlie the different sensitivities of these promoters to the DNA bending activity of TFAM (Fig. 5.9A, B). At LSP, the C-terminal tail is oriented away from the transcriptional start site, and a U-turn in the DNA is required to bring the C-terminal tail in contact with the transcriptional machinery (Fig. 5.9A). At HSP1, we directly con-

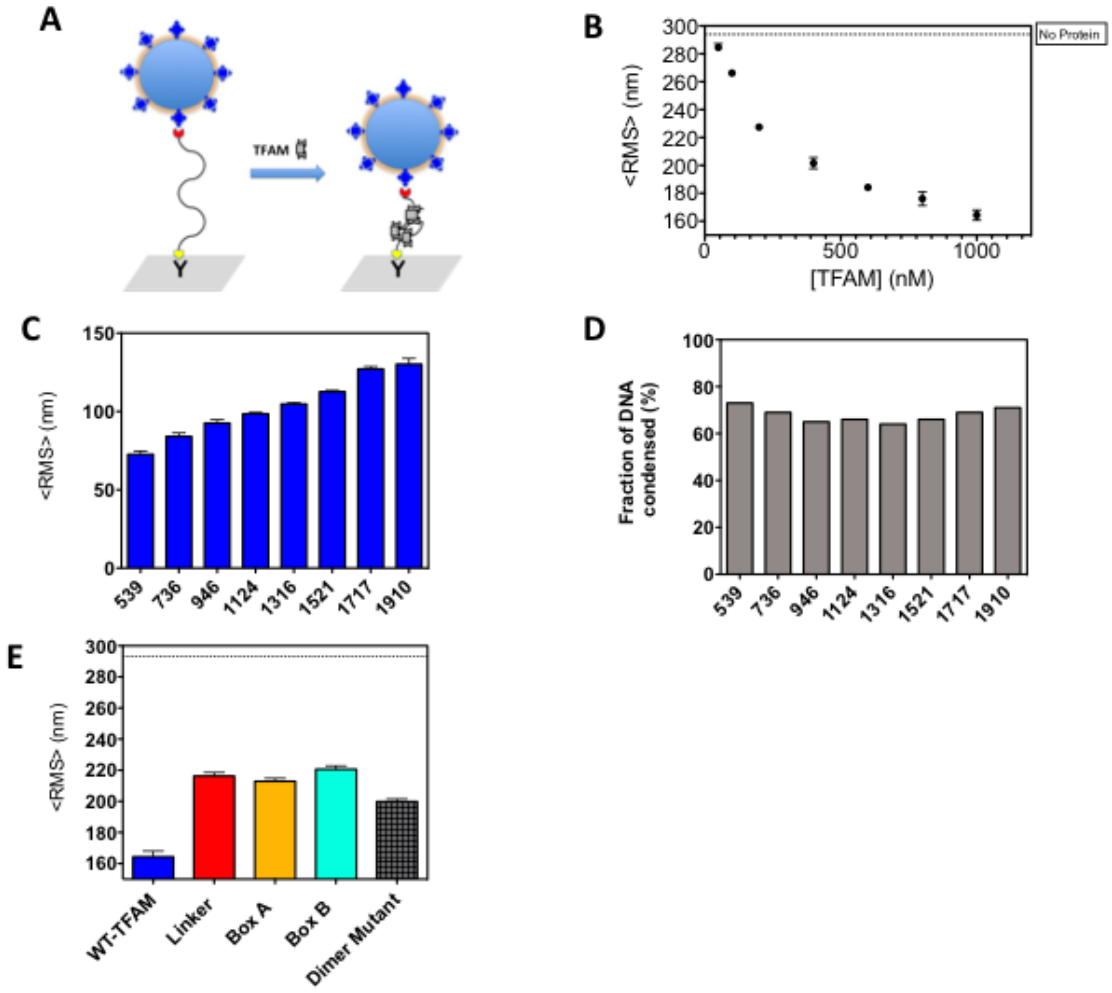

Figure 5.7: **Structural determinants of DNA compaction** (A) A schematic of the TPM assay. A single DNA molecule attaches the bead to the glass surface. Upon addition of TFAM, the contour length of the DNA molecule is reduced, causing a decrease in the bead's radius of motion. (B) Effect of increasing concentrations of TFAM on the DNA contour length. A 1910 bp DNA fragment was used. (C)DNA compaction by TFAM on DNA fragments of varying lengths. (D) Fractional shortening of DNA by TFAM as a function of DNA length. (E) DNA compaction by wildtype TFAM and mutants. A 1910 bp DNA fragment was used. (Note: See appendix 4.7 for experimental details, trajectories and distributions.)

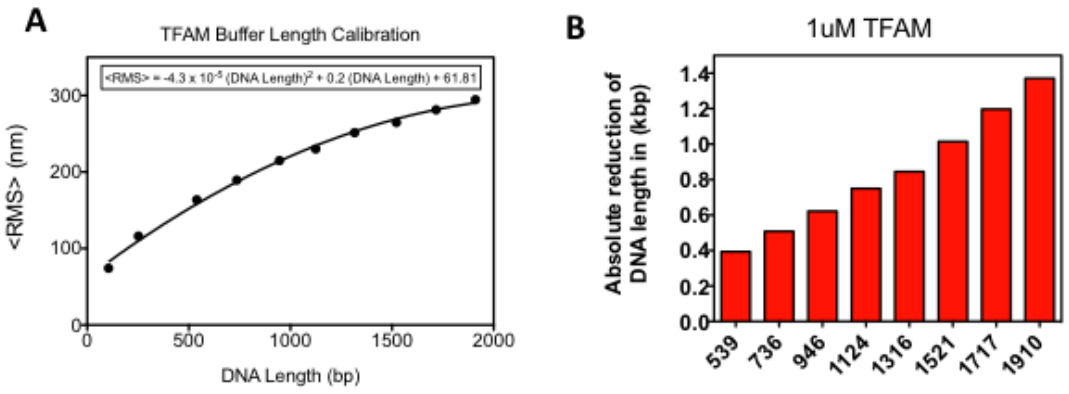

Figure 5.8: **DNA compaction in the TPM assay** (A) Calibration curve for linear DNA. The plot was generated by measuring RMS values for naked DNA of various lengths (250, 539, 736, 946, 1124, 1316, 1521, 1717, and 1910 bp). (B) Reduction of DNA length by TFAM for DNA of various lengths. For each length length, RMS values were measured with and without TFAM. The reduction in apparent DNA length was obtained from the calibration curve in (A).

firmed by bromine-labeling that the TFAM orientation is reversed. This new orientation places the C-terminal tail close to the transcriptional machinery and negates the need for DNA bending (Fig. 5.9B). This model is further supported by promoter engineering experiments, where we converted HSP1 into a DNA-bending-dependent promoter by reversing the TFAM orientation. Finally, we found that the four existing TFAM/DNA structures show dimerization of TFAM with an identical interface formed by HMG-box A/HMG-box A interactions. In our previous study, we ignored this interface as a potential crystal-packing artifact, but its presence in four independent structures makes this explanation unlikely. Using a FRET assay, we detected a TFAM-TFAM interaction that is strictly dependent on the presence of DNA. Analysis of the dimer mutant indicates that dimerization is not required for DNA bending or transcriptional activation at LSP or HSP1. However, the TFAM dimer mutant shows substantially reduced DNA compaction in a TPM assay. Taken together, these results suggest that specific structural features of TFAM are critical depending on the specific function. At LSP, DNA bending at the promoter is central to transcriptional activation, whereas at HSP1, this feature is dispensable. TFAM dimerization plays a key role in mtDNA compaction but not in transcriptional activation.

Our results suggest that two factors, DNA bending and TFAM dimerization are

involved in mtDNA compaction. In the TFAM/DNA structures, the formation of a U-turn compresses the 22 base-pair recognition sequence by 50% end-to-end. This extreme DNA bending requires spatial coordination of the tandem HMG-box domains by the helical linker. TFAM variants containing a single HMG-box domain or mutations in the helical linker show defective DNA bending and, in the TPM assay, reduced DNA compaction. Although DNA bending is clearly important, it alone cannot account for the 64-73% compaction that we observed in the TPM assay. Based on the behavior of the dimer mutant, this additional compaction appears to require TFAM dimerization. Given that the TPM assay measures the effective length of individual DNA tethers, this dimerization occurs between TFAM molecules bound to a single DNA fragment. TFAM dimerization in solution is only observed with DNA templates 200 base-pairs in length, probably because dimerization of TFAM molecules bound on the same DNA fragment would lead to DNA looping (Fig. 5.9C), a phenomenon that is constrained by the persistence length of DNA. In contrast, under the high protein/DNA concentrations used for crystallography, dimerization was observed between TFAM molecules bound to separate, short DNA fragments. It should be noted that the type of dimerization implied by our results is distinct from the dimerization suggested in previous studies. Some studies have suggested that TFAM binds to a single DNA binding site as a dimer [14, 20]. This mode of binding is unlikely, given the multiple crystal structures showing that the TFAM binding sites identified by DNAseI footprinting [10] are fully occupied by a single TFAM molecule. The idea that a TFAM monomer binds to DNA is further supported by static light scattering [8] and single molecule measurements [17].

These findings clarify the mechanism for mtDNA packaging by TFAM. One aspect involves iterations of mtDNA bending caused by coating of the genome by TFAM. Measurements in tissues [13, 21] and cells [22, 23] suggest that TFAM is abundant enough to fully coat the mitochondrial genome. Moreover, recombinant TFAM can coat nonspecific DNA *in vitro* with an occupancy of one molecule per 30 bp [17], close to the theoretical maximum based on a minimal TFAM binding site of 22 bp. The repeated application of DNA bending would cause significant shortening of the end-to-end length of a DNA molecule. This condensed DNA molecule can be further compacted by TFAM dimerization, which would result in looping of the DNA between bound sites. Dimerization can only occur between TFAM molecules spaced the appropriate distance apart. Interestingly, atomic force

microscopy images of DNA compacted *in vitro* by TFAM show formation of DNA loops consistent with this model [14].

5.5 Materials and Methods

5.5.1 TFAM purification and labeling

Full-length, mature human TFAM (43-246), TFAM lacking the last 9 residues (43-237), and mutants were expressed and purified as hexahistidine fusion proteins as described previously [8]. The hexahistidine tag was removed with thrombin cleavage. The proteins were further purified by gel filtration chromatography using a Hi-load Superdex 200 16/60 column (GE Healthcare) pre-equilibrated with running buffer (20 mM Tris-HCl, 300 mM NaCl, 1 mM DTT, pH 7.5) in an AKTA Purifier (Amersham). The protein was concentrated to 20 mg/ml and stored at $-80\,^{\circ}$C.To obtain fluorescently labeled TFAM, the protein was incubated with either Alexa Fluor 488 maleimide or Alexa Fluor 595 maleimide (Molecular Probes) as described [24] with modifications. The TFAM construct (43-237) contains only a single reactive cysteine at position 49. Briefly, the protein was dialyzed in buffer (50 mM HEPES, pH 7.5, 350 mM NaCl), diluted to 0.1 mM, and treated with 0.5 mM TCEP (tris(2-carboxyethyl)phosphine) for 90 minutes at $4\,^{\circ}$C to ensure reduction of cysteine residues. The labeling reaction was carried out using a six- fold excess of Alexa Fluor maleimide for 15 h at $4\,^{\circ}$C and stopped by 3 mM DTT. Unreacted dye was removed with a 15 cm desalting column (Sephadex G-25 Medium, GE Healthcare). The labeling efficiency was typically $\geq 95\%$.

5.5.2 Crystallization, data collection, and structure determination

The duplex HSP1 DNA was generated by mixing complementary oligonucleotides:

5- GGTT550GGTTCGGGGTATGGGGTT -3'
5- AACCCCATACCCCGAACCAACC -3'

The nonspecific mtDNA template was generated by mixing complementary oligonucleotides:

5- ATTCAACCAATAGCCCTGGCCG-3

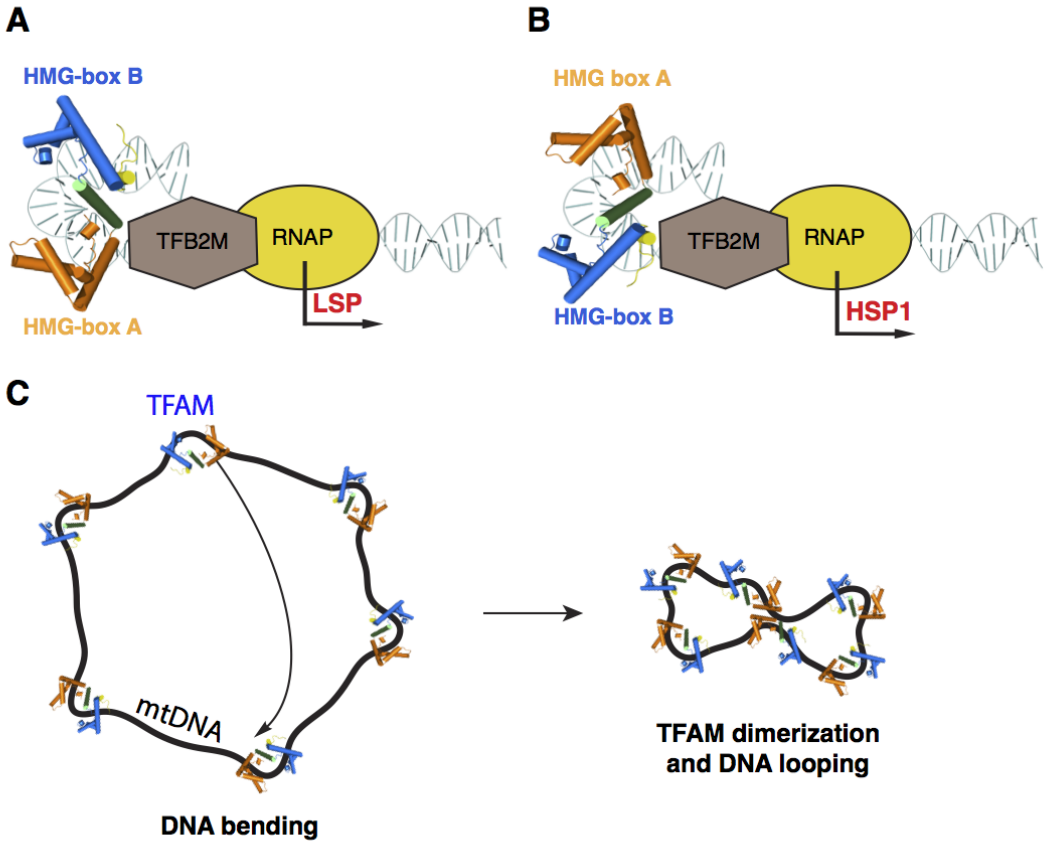

Figure 5.9: **Models of transcription activation by TFAM and mtDNA packaging** (A),(B) Comparison of TFAM function on LSP versus HSP1. When bound on LSP (A), TFAM is oriented with the HMB-box B domain binding the distal half-site. As a result, the U-turn in DNA is necessary to position the C-terminal tail towards the transcriptional machinery. The C- terminal tail is essential for transcriptional activation [9] and physically interacts with TFB2M [40], but additional interactions with mtRNA polymerase are also possible. When bound to HSP1 (B), TFAM is oriented with the HMG-box B domain binding the proximal half-site. The C-terminal tail is positioned close to the transcriptional machinery, and DNA bending is dispensable for transcriptional activation. (C) Model of mtDNA compaction. Upon binding to mtDNA, each TFAM molecule imposes a local U-turn. When TFAM coats mtDNA, the formation of multiple U-turns results in mtDNA compaction. In addition, TFAM monomers can dimerize through the HMG-box A domain. This interaction forms DNA loops, which further compact mtDNA in the nucleoid.

5- CGGCCAGGGCTATTGGTTGAAT-3.

For the bromine derivative of the HSP1 duplex, thymine 550 of HSP1 was substituted with bromo-uracil (5-GGT /i5Br-dU/GG TTC GGG GTA TGG GGT T-3) (Integrated DNA Technologies, San Diego) and annealed with its non-labeled complementary strand. To form the TFAM (43-237)/HSP1 or TFAM (43-237)/HSP1-Br complexes, 0.57 mM TFAM was mixed with 0.65 mM duplex DNA in a 1:1 volume ratio, incubated at room temperature for 30 min, and then on ice for 2 h. Crystallization trials by hanging drop-vapor diffusion at room temperature identified conditions (12% PEG8000, 22% glycerol, 0.04 M KH2PO4, pH 6.0) that yielded plate-shaped crystals. To form the TFAM (43-237)/nonspecific DNA complex, 0.57 mM TFAM was mixed with 0.65 mM duplex DNA in a 1:11 volume ratio, incubated at room temperature for 30 min, and then on ice for 3 h. For crystallization trials, the protein/DNA complex was mixed with reservoir solution in a 1.2: 0.8 complex/well solution volume ratio. The complex was crystallized by hanging drop-vapor diffusion at room temperature (9% PEG1500 and 18% glycerol), yielding plate-shaped crystals. Crystals were flash-frozen in liquid nitrogen, and diffraction data of native crystals were collected on beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL). Diffraction data for TFAM/HSP1-Br were collected at the Br-absorption peak. All data wereprocessed with IMOSFLM [25] or XDS [26] and merged using SCALA [27] as implemented in CCP4. The two structures were solved by molecular replacement using the coordinates of the TFAM/LSP crystal structure [19]. The electron density map revealed clear density for the protein and DNA. The bromine derivative of TFAM/HSP1 yielded anomalous signal that allowed unambiguous assignment of the DNA sequence. Manual model building was done in COOT [28], and refinement of the best solutions was carried out using PHENIX [29] with an initial round of rigid body refinement followed by a round of simulated annealing and individual B- factor refinement. After a few rounds of model adjustment and refinement with TLS obtained from the TLSMD server [30], the Rwork converged to 22.87% and Rfree to 26.63% for TFAM/HSP1, and 21.37% and 26.11% for TFAM/nonspecific DNA. The final models include residues 43-237 of TFAM, and all the nucleotides have excellent stereochemistry as assessed by MolProbity [31]. Analyses of the helical parameters of the DNA molecules were carried out using 3DNA [32].. Structural superimpositions and figures were prepared with Chimera 1.5.2 [33], COOT [28], PyMol [34], and 3DDART [35].

5.5.3 FRET experiments for DNA bending

To generate the LSP, HSP1, and nonspecific DNA templates, the following complementary oligonucleotides were annealed as described above:

LSP: 5/Cy3/-TGTTAGTTGGGGGGTGACTGTTAAAAGT
5/Cy5/-ACTTTTAACAGTCACCCCCCAACTAACA

HSP1: 5/Cy3/-GGTTGGTTCGGGGTATGGGGTT
5/Cy5/-AACCCCATACCCCGAACCAACC

Nonspecific: 5'/5Cy5/-ATTCAACCAATAGCCCTGGCCG
5'/5Cy3/-CGGCCAGGGCTATTGGTTGAAT

Annealing reactions were performed with a small excess of acceptor labeled DNA relative to donor labeled DNA (1.2:1 molar ratio) to ensure that all donor-labeled DNA had a paired acceptor DNA strand [36]. Measurements, data processing, and FRET calculations were performed as previously described [8, 36].

5.5.4 FRET measurements to detect dimerization

To detect dimerization of TFAM through the HMG-box A domain, Cys49 in HMG-box A was labeled with either Alexa Fluor 488 or Alexa Fluor 595 (see above). Fluorescence measurements were recorded on a FluoroMax-3 spectrofluorimeter (JobinYvonHoriba) at room temperature [20 mM TrisHCl (pH 7.5), 150 mM NaCl and 1 mM DTT]. 400 nM Alexa Fluor 595-labeled protein was added to 100 nM Alexa Fluor 488- labeled protein in the presence and absence of DNA. DNA tested included fragments of various lengths (100, 150, 200, 300, 400 bp) and circular plasmid (4 kb). When the experiment was performed with DNA, 10 nM of linear DNA, or 1 nM of circular plasmid DNA was added to the mixture containing both TFAM labeled with Alexa Fluor 488 and Alexa Fluor 594. The acceptor fluorescence signal could be chased off with excess non-labeled wild-type protein. The signal was monitored by excitation at 480 nm, and emission spectra were recorded from 490 to 690 nm. Slit widths for excitation and emission were 4 and 10 nm, respectively.

5.5.5 In vitro transcription reactions.

DNA fragments corresponding to LSP (positions 1-477) and HSP1 (positions 499-741) of human mtDNA were cloned into the pSP65 vector at the BamHI and SalI sites. EP1 was constructed by reversing the TFAM binding site at HSP1, and EP2 was made by subcloning the TFAM binding site from LSP into HSP1.

LSP: 5-A454AAT AATGTGTTAGTTGGGGGGTGACTGTT AAAA421
HSP1: 5-T528 GCT AACCCCATACCCCGAACCAACC AAAC557
EP1: 5-T528 GCT TGGTTCGGGGTATGGGGTTAGC AAAC557
EP2: 5-T528GCT AATGTGTTAGTTGGGGGGTGACTGTT AAAC557

The engineered promoters were synthesized by PCR with the appropriate oligonucleotides and cloned into the pSP65 vector at the BamHI and SalI sites. After digestion with BamHI for LSP and SalI for HSP1 derivatives, the linearized plasmids were used as templates in a transcriptional run-off assay as described [8]. In brief, transcription reactions were performed in a total volume of 20 μL [10 mM HEPES (pH 7.5), 10 mM MgCl2, 1 mM DTT, 100 μgmL BSA, 400 μM ATP, 150 μM rCTP, 150 μM rGTP, 15 μM rUTP (Promega), 0.2 μM [α-32P] rUTP (3,000 Cimmol, Perkin Elmer), 7.5 nM template DNA, and 40 units of RNaseOut (Invitrogen)]. Reactions were carried out at 33 °C by incubating the template DNA with buffer for 5 min, then subsequently adding TFAM (100 nM), TFB2M (30 nM, Enzymax), and POLRMT (30 nM, Enzymax), with 1 min incubations between each addition. After addition of rNTPs, the reaction was incubated for 3 h at 33 °C and stopped by addition of 25 μL of stop buffer (80% formamide, 10 mM EDTA, pH 8.0, 0.025% xylene cyanol, 0.025% bromophenol blue). Samples were heated to 90 °C for 5 min and separated on 5% polyacrylamide gels containing 8 M urea in 1xTBE. The gels were fixed in 7% acetic acid, dried, and exposed to a phosphorimager screen. The data were collected on a Storm 880 phosphorimager (Molecular Dynamics) and quantified using ImageQuant 5.2 Software.

5.5.6 Tethered particle motion (TPM) assays.

Non-sequence specific DNA molecules of different lengths (539, 736, 946, 1124, 1316, 1521, 1717, and 1910 bp) were assembled in a flow cell between a glass surface and a bead.

These DNA molecules were immobilized by attaching the 5-digoxigenin-labeled DNA to a glass surface coated with anti-digoxigenin (Roche Diagnostics). The other end of the DNA was attached to a 490-nm streptavidin-coated polystyrene bead (Bangs Laboratories) via a 5-biotin label. The flow cell surface was passivated with acetylated-BSA and casein to prevent the DNA, beads and protein of interest from sticking during the experiments. Flow cell preparation, data acquisition and analysis were performed as previously described [37]. 20 single DNA tethers were measured simultaneously in a buffer containing 20 mM TrisHCl, pH 7.5, 1 mM DTT, 150 mM NaCl and 100 μg/ml A-BSA.

5.6 Appendix 1: TPM experimental details, trajectories and distributions

5.6.1 TPM flow cell and DNA-tethered bead assembly

The experimental procedure for the TPM-TFAM experiments shown in Fig. 5.7B, C, and E begins by first drilling 12 holes into a pre-cleaned glass slide with a 1.5mm diamond tipped drill bit (CRLaurence) to create inputs and outputs for six channels. The drilled glass slide (Corning) and a partner coverslip (Fisher) are sonicated (Aquasonic-Model 75T) in 100% EtOH for 30 min, rinsed with ddH_2O, sonicated in 1M KOH for 30 min, rinsed with ddH_2O, sonicated in 5% HCl, rinsed with ddH_2O, and functionalized with epoxysilane (116mL Isopropanol (IPA), 1ml ddH_2O, 3ml of 3-glycidoxypropyldimethoxymethylsilane (Sigma) and 120 μL N,N-Dimethylbenzylamine (Sigma)) at RT for 1 h and 30 min. The drilled glass slides and coverslips were washed in IPA, dried with N2 gas and baked at 100C for 30 min. Next 0.02 in. ID/0.06 in. OD tygon microbore tubing (Cole-Parmer) is placed in each of the 12 holes of the drilled glass slide, designating an input and output for all six channels. Next five-minute epoxy (Devcon) is applied at the tubing-glass interface, and a piece of double-sided tape (0.12mm thick, Grace Bio-Labs) with six holes designating the channels of the flow cell is placed on the drilled glass slide and is sandwiched between a coverslip. After the flow cell is assembled polyclonal sheep anti-digoxigenin (Roche) is added to each of the six channels to a final concentration of 40ng/μL. The anti-digoxigenin (anti-dig) solutiion is incubated overnight at RT to allow covalent coupling of anti-dig to epoxy groups within the channels of the flow cell. Next, TPM assembly buffer (TPM-AB: 20mM Tris-Acetic Acid pH 8.0, 130mM KCl, 1mM MgCl2, 0.1mM DTT, 0.1mM EDTA, 20ug/ml Aceylated-BSA and 3mg/ml Biotin-free Casein) is added to the flow cell's channels and incubated for 1hr in order to flush out uncoupled anti-digoxigenin and quench active epoxy groups. Then biotin and digoxigenin labeled DNA is added to each channel to a final concentration of 5pg/μL and incubated for 20 minutes. The excess DNA is washed away with TPM-AB and 0.49μm streptavidin coated polystyrene beads (Bangs Laboratories) diluted in TPM-AB to a final concentration 3 X 10^{10} beads/mL are added to each channel and incubated for 10 min. The excess beads are washed away with TPM-AB and TFAM binding buffer (20 mM TrisHCl, pH 7.5, 1 mM DTT, 150 mM NaCl and 100 μg/ml A-BSA)

completing the assembly of DNA-tethered beads.

5.6.2 TPM Acquisition and Analysis

The DNA tethered beads were imaged using brightfield microscopy (because the polystyrene beads scatter light) on inverted Olympus IX71 microscopes with either a 100x oil objective, or a 60x oil objective with a 1.6x magnifier. A Basler A602f camera was used to acquire images at 60Hz; however every other frame was dropped for a final frame rate of 30Hz (Note: This was implemented by a previous student to compare results imaged using a CCD camera with a frame rate of 30Hz. In these studies there is no particular reason we acquired at 30Hz). We track the DNA-tethered beads by cross-correlating each frame with the initial frame in a time series for each bead. This generates "raw" x and y positions of each bead's excursion, relative to its anchor point. Drift was corrected from these raw data by subtracting the results of a low-pass first-order Butterworth filter with a cutoff frequency $f_{cB} = 0.05$ Hz. The root-mean-square (RMS) motion was obtained by applying a Gaussian filter with a -3 dB frequency of $f_{cG} = 0.0326$ Hz corresponding to a 4 second or 120 frame standard deviation of the filter, to the mean-squared displacement of the data ($\overrightarrow{x}^2 + \overrightarrow{y}^2$). The RMS motion of the bead is the square root of the result of the convolution of ($\overrightarrow{x}^2 + \overrightarrow{y}^2$) and the Gaussian filter.

To define a "good" DNA tethered bead we used symmetry-of-motion and length-of-motion checks used previously in [Han L et al 2009]. We then track the good beads in buffer lacking TFAM for at least 350 sec, then we stop tracking the beads, and add a given amount of TFAM or a TFAM mutant, and continue tracking the same beads. This scheme allows for precision measurements of changes to the RMS motion of DNA-tethered beads by TFAM. "Bad" beads who passed the initial checks are discarded if the beads stick to the surface and if the beads initial RMS is not within error to the previously measured RMS for a particular DNA length.

5.6.3 TPM trajectories and distributions from Fig. 5.7B, C and E

Fig 5.7B showed that TFAM reduces the RMS of DNA-tethered beads in a concentration dependent manner. TFAM was added to a channel with a non-specific 1910bp DNA-tethered beads at concentrations of 50nM, 100nM, 200nM, 400nM, 600nM , 800nM and 1μM. TFAM's dissociation constant is 7nM, however a minimal TFAM concentration of

50nM is required to measure a shift in the RMS motion. This means there are multiple TFAMs per DNA tethered bead which is a why all the trajectories and distributions over the concentration range are displaced the entire time, showing no transitioning between states. The 1910bp DNA substrate was amplified from the pUC19-NoRSS. The trajectories and normalized probability distributions from the titration are shown in Fig 5.10 (A-G). A few of the trajectories show increasing or decreasing spikes in the RMS. The increasing spikes are the result of beads in solution passing through the field which increases the apparent radius of motion which in turn increases the RMS motion. The decreasing spikes are transient sticking of the DNA-tethered bead, which reduces the RMS motion. Fig 5.7C

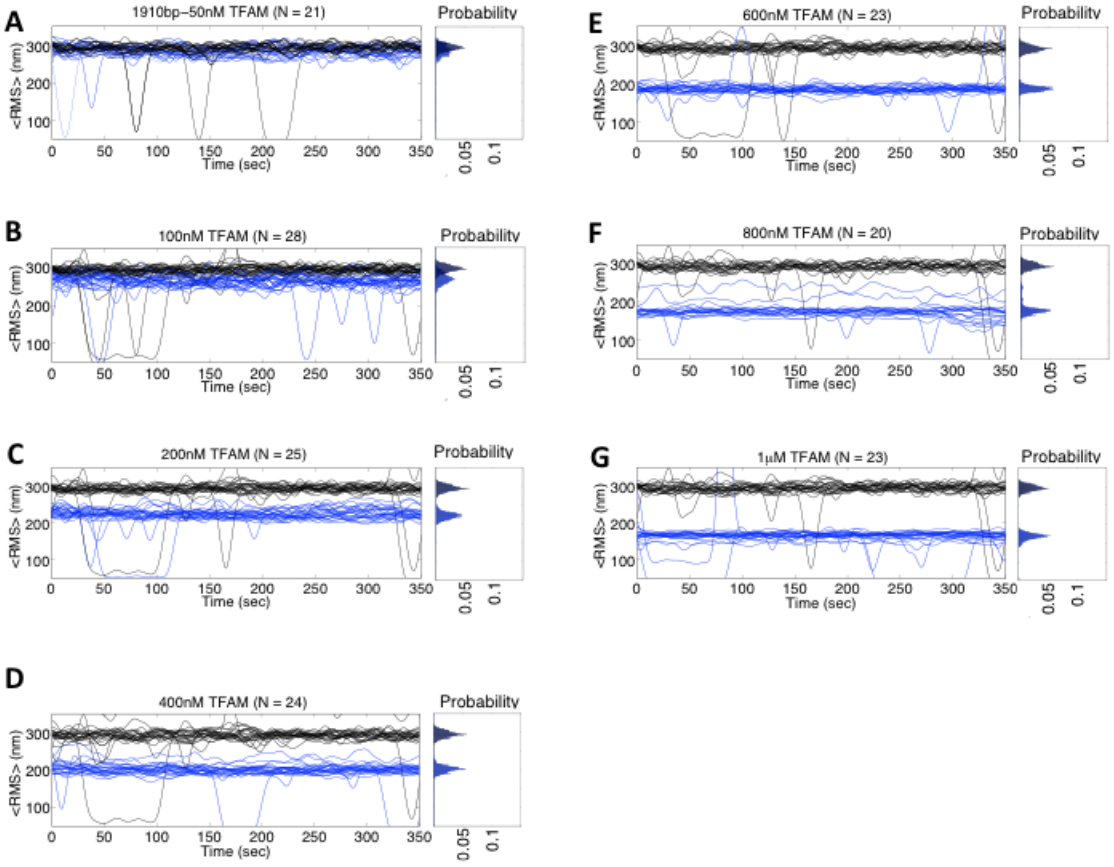

Figure 5.10: **TFAM compacts DNA in a concentration dependent manner.** TFAM was titrated over a concentration range of 50nM-1μM in the presence of a 1910bp DNA substrate. (A-G) are the trajectories that went into the data points in Fig. 5.7B. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before TFAM is added are black and the trajectories after TFAM is added are blue.

showed the absolute change in the RMS increased with increasing DNA length. TFAM was added to channels with non-specific DNA of various lengths (539bp, 736bp, 946bp, 1124bp, 1316bp, 1521bp, 1717bp or 1910bp) at 1μM. The various DNA substrates were derived from pUC19-NoRSS. The trajectories and normalized probability distributions from the length calibration are shown in Fig 5.11(A-H).

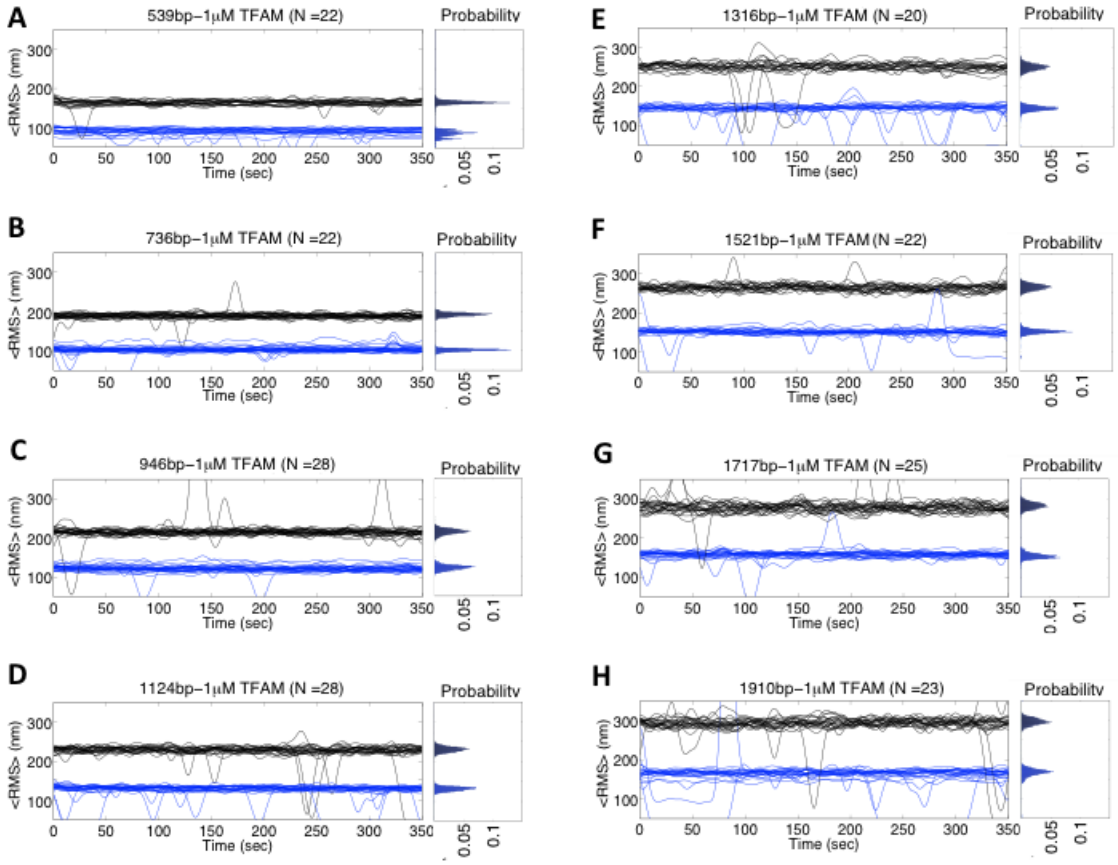

Figure 5.11: **TFAM compacts DNA of various DNA lengths.** TFAM was fixed at 1μM and added to DNA substrates of lengths 539bp, 736bp, 946bp, 1124bp, 1316bp, 1521bp, 1717bp and 1910bp. (A)-(H) are the trajectories that went into the data points in Fig. 5.7C. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before TFAM is added are black and the trajectories after TFAM is added are blue.

Fig 5.7E showed that TFAM mutants were defective at compacting DNA relative to WT-TFAM. Various TFAM mutants were added to channels with a 1910bp substrate. The trajectories and normalized probability distributions from TFAM mutant data are shown in Fig 5.12(A-E)

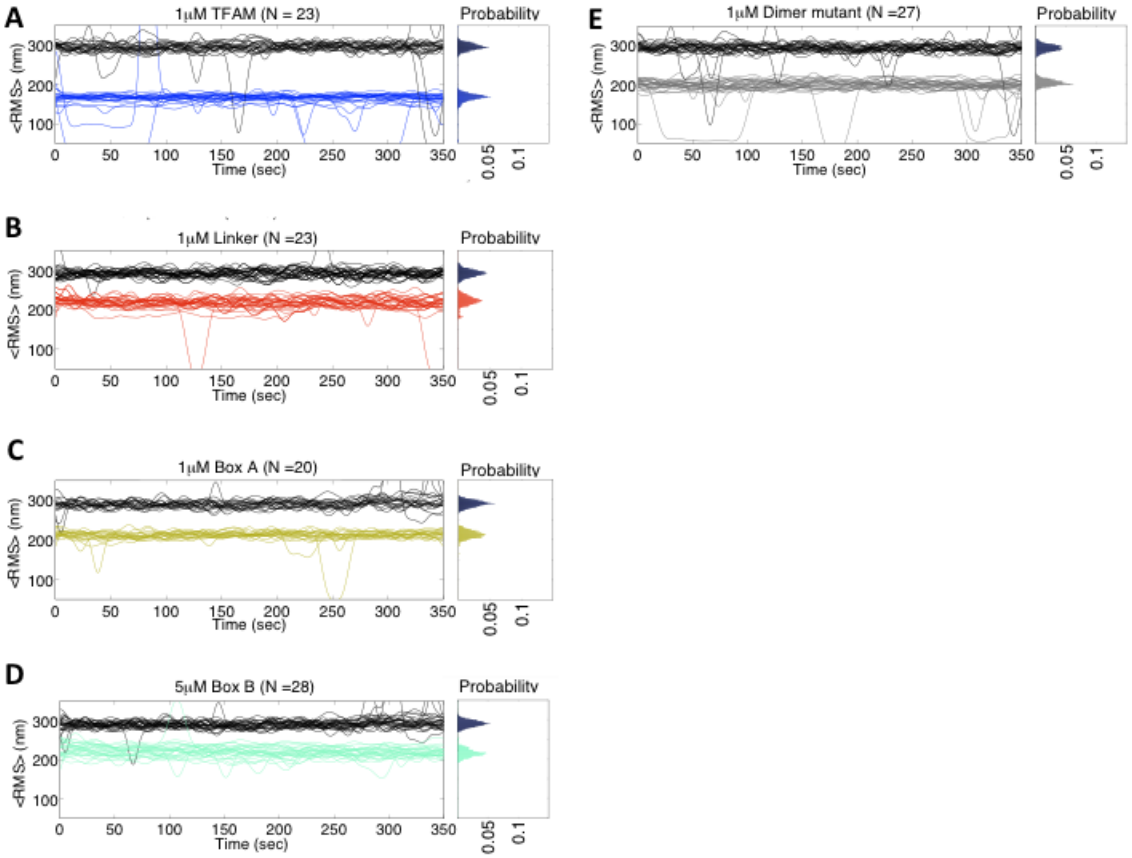

Figure 5.12: **TFAM mutants are defective at DNA compaction.** TFAM mutants linker, boxA, boxB and dimer mutant were added to a 1910bp DNA substrate. (A-E) are the trajectories that went into the data points in Fig. 5.7E. DNA-tethered beads were tracked for at least 350 secs at 30Hz. Each trajectory is averaged over 120 frames or 4 sec. A composite normalized probability distribution of all DNA-tethered beads is plotted adjacent to the trajectories. The trajectories before linker, boxA, boxB and dimer mutants are added are black and the trajectories after the linker, boxA, boxB and dimer mutants are added are red, orange, cyan and dark gray respectively.

5.7 References

[1]. Bonawitz, N.D.*et al.* (2006). Initiation and beyond: multiple functions of the human mitochondrial transcription machinery. *Mol.Cell* 24, 813-825.

[2]. Campbell, C.T., *et al.* (2012). Mitochondrial transcription factor A regulates mitochondrial transcription initiation, DNA packaging, and genome copy number. Biochim Biophys Acta 1819, 921-9.

[3] Falkenberg, M., *et al.*(2007). DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem* 76, 679-99.

[4.] Rubio-Cosials, A. and M. Sola (2013). U-turn DNA bending by human mitochondrial transcription factor A. *Curr Opin Struct Biol* 23, 116-24.

[5]. Falkenberg, M., *et al.* (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat.Genet.* 31, 289-294.

[6]. Shi, Y., *et al.* (2012). Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *PNAS* 109, 16510-16515.

[7]. Larsson, N.G., *et al.* (1998). Mitochondrial transcription factor A is necessary for mtDNA maintenance and embryogenesis in mice. *Nat Genet* 18, 231-6.

[8]. Ngo, H.B., J.T. Kaiser, and D.C. Chan (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat.Struct.Mol.Biol.* 18, 1290-1296.

[9]. Dairaghi, D.J., *et al.* (1995). Addition of a 29 residue carboxyl- terminal tail converts a simple HMG box-containing protein into a transcriptional activator. *J.Mol.Biol.* 249, 11-28.

[10]. Fisher, R.P., *et al.* (1987). Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. *Cell* 50, 247-58.

[11]. Bogenhagen, D.F., *et al.* (2008). The layered structure of human mitochondrial DNA nucleoids. *J Biol Chem* 283, 3665-75.

[12]. Bogenhagen, D.F., *et al.* (2003). Protein components of mitochondrial DNA nucleoids in higher eukaryotes. *Mol Cell Proteomics* 2, 1205-16.

[13]. Alam, T.I., *et al.* (2003). Human mitochondrial DNA is packaged with TFAM. *NAR* 31, 16‘40-5. PMCID: 152855.

[14]. Kaufman, B.A., *et al.* (2007). The mitochondrial transcription factor TFAM coordinates the assembly of multiple DNA molecules into nucleoid-like structures. *Mol Biol Cell* 18, 3225-36. PMCID: 1951767.

[15]. Mouw, K.W. and P.A. Rice (2007). Shaping the Borrelia burgdorferi genome: crystal structure and binding properties of the DNA-bending protein Hbb. *Mol Microbiol* 63, 1319-30.

[16]. Rice, P.A., *et al.* (1996). Crystal structure of an IHF- DNA complex: a protein-induced DNA U-turn. *Cell* 87, 1295-306.

[17]. Farge, G., *et al.* (2012). Protein sliding and DNA denaturation are essential for DNA organization by human mitochondrial transcription factor A. *Nat Commun* 3, 1013.

[18]. Fisher, R.P. and D.A. Clayton (1988). Purification and characterization of human mitochondrial transcription factor 1. *Mol Cell Biol* 8, 3496-509. PMCID: 363587.

[19]. Rubio-Cosials, A., *et al.* (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat.Struct.Mol.Biol.* 18, 1281-1289.

[20]. Gangelhoff, T.A., *et al.* (2009). Structural analysis and DNA binding of the HMG domains of the human mitochondrial transcription factor A. *NAR* 37, 3153-64. PMCID: 2691818.

[21]. Ekstrand, M.I., *et al.* (2004). Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum Mol Genet* 13, 935-44.

[22]. Kukat, C., *et al.* (2011). Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *PNAS* 108, 13534-9. PMCID: 3158146.

[23]. Takamatsu, C., S. *et al.* (2002). Regulation of mitochondrial D-loops by transcription factor A and single-stranded DNA-binding protein. *EMBO Rep* 3, 451-6. PMCID: 1084112.

[24]. Zhang, X., S. *et al.* (2008). Demonstration of a multistep mechanism for assembly of the SRP x SRP receptor complex: implications for the catalytic role of SRP RNA. *J.Mol.Biol.* 381, 581-593.

[25]. Leslie, A.G. (1999). Integration of macromolecular diffraction data. *Acta Crystallogr.D.Biol.Crystallogr.* 55, 1696-1702.

[26]. Kabsch, W. (2010). XDS. *Acta Crystallogr.D.Biol.Crystallogr.* 66, 125-132.

[27]. Evans, P.R., in Proceedings of the CCP4 Study Weekend. Data Collection and

Processing, L. Sawyer, N. Isaacs, and S. Bailey, Editors. 1993, Warrington: Daresbury Laboratory. p. 114-122.

[28]. Emsley, P. and K. Cowtan (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr.D.Biol.Crystallogr.* 60, 2126-2132.

[29]. Adams, P.D., *et al.*(2011). The Phenix software for automated determination of macromolecular structures. *Methods* 55, 94-106.

[30]. Painter, J. and E.A. Meritt (2006). TLSMD web server for the generation of multi-group TLS models. *J. Appl. Cryst.* 39, 109-111.

[31]. Davis, I.W., *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *NAR.* 35, W375-W383.

[32]. Lu, X.J. and W.K. Olson (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *NAR*, 5108-21. PMCID: 212791.

[33]. Pettersen, E.F., *et al.* (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-12.

[34]. Delano, W.L., The PyMOL Molecular Graphics System. 2002, San Carlos, CA: Delano Scientific.

[35]. van, D.M. and A.M. Bonvin (2009). 3D-DART: a DNA structure modelling server. *NAR* 37, W235-W239.

[36]. Hieb, A.R., *et al.* (2007). TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. *J Mol Biol* 372, 619-32.

[37]. Han, L., *et al.* (2009). Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One* 4, e5621. PMCID: 2682762.

[38]. Finzi, L. and J. Gelles (1995). Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules. *Science* 267, 378-80.

[39]. Johnson, S., *et al.* (2012). Sequence dependence of transcription factor-mediated DNA looping. *NAR* 40, 7728-7738.

[40]. McCulloch, V. and G.S. Shadel (2003). Human mitochondrial transcription factor B1 interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. *Mol Cell Biol* 23, 5816-24. PMCID: 166325.

Chapter 6

Conclusion and Future Directions

Hopefully this thesis has convinced you the genome is a physical object that can be manipulated with precision by protein effectors. The effectors can traverse the genome with base pair precision and carry out essential biological functions for an organism's development. However, we do not have an intuition for how most of these machines work with precision because the bulk biochemical and biophysical techniques previously used to understand these processes are confounded by ensemble averaging. Also the difficulty of isolating some of these protein machines and producing enough material to assay their function has stifled our pursuit to understand these processes.

6.1 V(D)J recombination

Here, we developed a new approach to purify large amounts of RAG proteins, which assemble antigen receptors. We accomplished this by producing RAG proteins from a new expression vector in human cells that are grown in suspension. The expression vector has the EBV origin of replication that is enhanced by EBNA1. Our work suggests that this is the mechanism at work; however, to definitively show that the increase in expression is a result of EBV and EBNA1, we must perform loss and gain of function experiments. One experiment would swap the EBV origin in the pTT5 vector for the SV40 origin in pcDNA1. The expected result would be an increase in RAG expression from 293-E cells transfected with pcDNA1 carrying an EBV origin. Currently, the precise mechanism responsible for the boost in expression is unknown, but the addition protein yield now allows us to pursue herculean experiments, for example, crystallizing the RAG proteins.

In addition, we developed single molecule assays to probe the action of the RAG

proteins. In these experiments we developed methods to measure the dissociation constants of the RAG recombinase with a statistical mechanical model that used Gaussian fits or kinetic theory to determine the probability of being bound to a 12RSS or 23RSS. To our knowledge the dissociation constant of the 23RSS has never been measured. This single molecule technique positions us to measure the binding constant of all non-consensus RSSs for any jawed-vertebrate. It would be very interesting to use this approach to try and solve the inverse problem for antigen-receptor assembly based solely on RSSs. The idea would be to determine if the affinity of an RSS site is a reflection of V, D, or J gene segment usage in an immune repertoire. The zebrafish would be great candidate for a study of this nature because 3D matrices have been generated showing the relative abundance of particular V, D, and J gene segments. We also used kinetic theory to extract a mean dwell time for the RAG1/2c at the site of an RSS. This value is completely new to the literature and answers the question, how long does a RAG complex sit at a 12 or 23RSS. In the future it would be a great interest to determine the dissociation constant and dwell times for the full-length RAG proteins.

In addition, we capture cooperative binding of RAG1/2c and HMGB1 at an RSS, and measure its dissociation constant. We also revealed that the HMGB1 enhances RAG1/2c-mediated DNA compaction on both the 12RSS and 23RSS, it was previously thought to only enhance the 23RSS. The precise number of HMGB1 molecules involved in forming a RAG-HMGB1-RSS complex is unknown, we have made an HMGB1 mutant with a single cysteine for fluorescent labeling to count sthe number of HMGB1 molecules. Since we were able to capture cooperative action of a co-factor of the RAG recombinase, we can use our single molecule platform to probe other proteins that have been implicated in the process, e.g. non-homologous end joining proteins, sequence-specific bending proteins and modified histones in future studies. We also directly observed paired complex formation, showed that it was reversible and that its dwell time was about 5 minutes. This is the first observation of paired complex formation dynamics in real-time, and its reversibility suggests the RAG recombInase samples variable region gene segments during antigen receptor gene assembly. The short life-time of the paired complex also has implications for the fidelity of the process. There are cryptic RSSs every 1kb in the genome and it begs the question why does the RAG recombinase completely rearrange the genome using cryptic RSSs? The short life-time of the paired complex using the most optimal RSSs and since a 1/4 of paired complexes

result in cleavage suggests the cleavage of cryptic RSSs is highly unlikely and provides an explanation for the why the RAGs do not carryout significant off-target rearrangement at cryptic RSSs. In future studies it would be very interesting to create DNA substrates with precise distances between every D and J gene segment, V and D segment and V and J segment (once D's are absent).

We also observed 12/23 rule-regulated hairpin production and transitions in the DNA tether length indicative of a paired complex but shorter than 10 secs. To increase our temporal resolution we can employ single molecule FRET to capture paired complex.

6.2 Genome Packaging

During single molecule studies of HMGB1 we found that the proteins compact DNA independent of contour length, and the magnitude of compaction could vary depending on the divalent cation. These findings contribute to our understanding of HMGB1-mediated compaction, but the number of HMGB1 molecules is uncertain. We could take a theoretical and experimental approach to get a feeling for the numbers. Using monte carlo simulations, we could generate a DNA molecule and compute its <RMS>, which is the same readout of our TPM results. We can then dock on HMGB1 molecules that each provide a $67\,^{\circ}\pm1.3\,^{\circ}$kink and completely saturate the DNA molecule. Then we can remove HMGB1 molecules one at a time and compute <RMS> for N-1 and repeat. To directly measure the number of HMGB1 molecules, we can use the HMGB1 construct with a single cysteine added to DNA tethers of different contour lengths at the same HMGB1 concentration.

Our study on TFAM provided novel crystal structures of the protein bound to distinct DNA variants and showed that complete DNA compaction required TFAM's dimerization interface. The contribution of the dimerzation interface suggests that TFAM loops the DNA, but in TPM we do not observe loop formation, although that does not mean looping is not occurring. One approach to resolve TFAM mediated loops is to use magnetic tweezers and pull on the DNA tethers. If the steps generated by removing individual TFAM molecules contain a step or steps larger then most, then that may be a signature of loop being disrupted.